

PR 01-SEP-1999; 99WO-US20111.
 PR 25-OCT-1999; 99US-0162505.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 16-DEC-1999; 99WO-US30095.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00376.

XX (GENTH) GENENTECH INC.

XX Baker KP, Botstein D, Desnoyers L, Eaton DI, Ferrara N, Fong S;
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CU, Gurney AL, Hillan KJ;
 PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
 PI Matanabe CK, Williams PM, Wood WI;
 XX WPI; 2001-071395/08.

XX Secreted and transmembrane proteins and nucleic acids designated PRO,
 PT useful as hybridization probes, in chromosome and gene mapping and gene
 therapy -

PS Claim 1; Fig 66; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.
 CC These proteins and the DNA encoding them may be used as hybridization
 CC probes, in chromosome and gene mapping and in the generation of
 CC anti-sense RNA and DNA. They may also be used to generate either
 CC transgenic animals or knockout animals which are in turn useful for
 CC development and screening of therapeutically useful reagents.
 CC The nucleic acids may also be used in gene therapy.

XX Sequence 331 AA;

Query Match 100.0%; Score 1695; DB 22; Length 331;
 Best Local Similarity 100.0%; Pred. No. 4.5e-166;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLPLSALGTGAGAAVLLKDYVTGACPSKATIPGKTIVYTANTGIGKOTALELR 60
 DB 1 MSRYLPLSALGTGAGAAVLLKDYVTGACPSKATIPGKTIVYTANTGIGKOTALELR 60
 QY 61 RGNITILACRMEKCEAAAKDIRGTINHNHVNARHIDLASISIREPAKIIIEEERVDI 120
 DB 61 RGNITILACRMEKCEAAAKDIRGTINHNHVNARHIDLASISIREPAKIIIEEERVDI 120
 QY 121 LINNAGVWRCPWTTEDGEMQGVNHLGHFLTNLLDLKASAPRIINLSLAIVAG 180
 DB 121 LINNAGVWRCPWTTEDGEMQGVNHLGHFLTNLLDLKASAPRIINLSLAIVAG 180
 QY 181 HIDPDILNQTRKRYNTKAAVCCSKLAIVFTKELSRRLQSGGVTVNALHPGVANTEIGRH 240
 DB 181 HIDPDILNQTRKRYNTKAAVCCSKLAIVFTKELSRRLQSGGVTVNALHPGVANTEIGRH 240
 QY 241 TGIHGSTFSGSTLGPFIWLIWKSPELAQPSSTYLAABEELADVSGKYEDGKKAPAPAEA 300
 DB 241 TGIHGSTFSGSTLGPFIWLIWKSPELAQPSSTYLAABEELADVSGKYEDGKKAPAPAEA 300
 QY 301 EDEEVARRLMAESARLVGLEAPSVREQPLPR 331
 DB 301 EDEEVARRLMAESARLVGLEAPSVREQPLPR 331

RESULT 4
 ID AAB84366 standard; Protein; 331 AA.

XX AAB84366;
 XX 22-AUG-2001 (first entry)
 XX Amino acid sequence of human alcohol dehydrogenase 21676.
 XX Human, alcohol dehydrogenase; colon disorder; brain disorder;

KM skin disorder; heart disorder; blood vessel disorder; kidney disorder;
 KM prostate disorder; skeletal muscle disorder; ovary disorder;
 KM testis disorder; epididymis disorder; spleen disorder; lung disorder;
 KM liver disorder; uterus disorder; endometrium disorder; T-cell disorder;
 KM red cell disorder; thymus disorder; B cell disorder; breast disorder;
 KM thyroid disorder; pancreas disorder; small intestine disorder;
 KM reduced platelet number disorder; precursor T cell neoplasm.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Modified-site 12..17 /note= "putative N-myristoylation site"
 XX Modified-site 28..33 /note= "putative N-myristoylation site"
 XX Modified-site 45..50 /note= "putative N-myristoylation site"
 XX Modified-site 54..57 /note= "putative N-myristoylation site"
 XX Modified-site 100..102 /note= "putative casein kinase II phosphorylation site"
 XX Modified-site 103..105 /note= "putative protein kinase C phosphorylation site"
 XX Modified-site 103..105 /note= "putative protein kinase C phosphorylation site"
 XX Modified-site 103..106 /note= "putative casein kinase II phosphorylation site"
 XX Modified-site 134..137 /note= "putative casein kinase II phosphorylation site"
 XX Modified-site 171..174 /note= "putative N-glycosylation site"
 XX Modified-site 191..193 /note= "putative protein kinase C phosphorylation site"
 XX Modified-site 215..217 /note= "putative protein kinase C phosphorylation site"
 XX Modified-site 220..225 /note= "putative N-myristoylation site"
 XX Modified-site 231..236 /note= "putative N-myristoylation site"
 XX Modified-site 242..247 /note= "putative N-myristoylation site"
 XX Modified-site 284..186 /note= "putative N-myristoylation site"
 XX Modified-site 313..315 /note= "putative protein kinase C phosphorylation site"
 XX Modified-site 323..325 /note= "putative protein kinase C phosphorylation site"
 XX Modified-site 323..326 /note= "putative casein kinase II phosphorylation site"
 XX WO200144446-A2.
 XX 21-JUN-2001.
 XX 15-DEC-2000; 2000WO-US33873.
 XX 15-DEC-1999; 99US-0464039.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX Meyers R;
 XX WPI; 2001-390244/41.
 XX N-PSDB; AAH25133.
 XX Novel human alcohol dehydrogenase proteins, 21612, 21615, 21620, 21676,
 PT 33756, useful for treating porriasia, tropical sprue, pancreatitis,
 PT goiter, osteomalacia, endometriosis, angina pectoris, embolism -
 XX Claim 9; Fig 11; 156pp; English.
 XX AAB84364-68 represent human alcohol dehydrogenase proteins, designated
 CC 21620, 33756, 21676, 21612 and 21615, respectively. Alcohol dehydrogenase
 CC polynucleotides and polypeptides are useful for treatment and diagnosis
 CC of disorders mediated by or related to alcohol dehydrogenase. They can

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2004, 11:26:06 ; Search time 52 seconds
(without alignments)
1010.356 Million cell updates/sec

Title: US-10-007-194a-116

Perfect score: 1695

Sequence: 1 MSRYLLPLSLGTVAGAAVL.....ESARLVGLAPSVREQPLR 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1695	100.0	331	21	AA199372
2	1695	100.0	331	22	AA605174
3	1695	100.0	331	22	AA866121
4	1691	99.8	331	22	AA884366
5	1691	99.8	331	23	AAU76232
6	1688	99.6	331	22	AA88356
7	1340	79.1	260	22	AA884365
8	1340	79.1	260	23	AAU76221
9	1332	78.6	260	22	AA40387

10	1136	67.0	240	22	AAE03299	Human gene 8 encod
11	1136	67.0	240	22	AAE01684	Human gene 13 enco
12	1136	67.0	240	23	ABG63953	Human albumin fusi
13	1136	67.0	240	23	ABG64446	Human albumin fusi
14	1136	67.0	273	21	AA857079	Human prostate can
15	1136	67.0	273	22	AAE03357	Human gene 8 enco
16	1014	59.8	197	22	AAE01756	Human gene 13 enco
17	947	55.9	290	22	AA442173	Human polypeptide
18	794	46.8	184	22	ABG15412	Novel human diagno
19	789.5	46.6	286	22	ABE58831	Drosophila melanc
20	767.5	45.3	300	22	ABE58826	Drosophila melanc
21	739.5	43.6	318	21	AA812158	Hydrophobic domain
22	739.5	43.6	318	21	AA82014	Human immunogenic
23	739.5	43.6	318	22	AAU69774	Human prostate CDN
24	739.5	43.6	318	22	AAU63777	Human polypeptide
25	739.5	43.6	318	22	AAU01129	Human prostate-spe
26	739.5	43.6	318	22	AAU699014	Human prostate-spe
27	739.5	43.6	318	22	AAU74812	Prostate tumour an
28	739.5	43.6	318	23	ABG65558	Human short chain
29	739.5	43.6	318	23	ABE58234	Human P509S protei
30	739.5	43.6	318	24	ABU71655	Prostate cancer as
31	739.5	43.6	333	20	AAU48616	Human breast tumo
32	738.5	43.6	314	21	AA836900	Human ARSBR1 prote
33	737.5	43.5	309	22	ABE58829	Drosophila melanc
34	737.5	43.5	316	23	AAU14438	Human drug metabol
35	732	43.2	321	23	ABE69258	Human polypeptide
36	726	42.8	336	20	AAU41730	Human PRO1072 prot
37	726	42.8	336	21	AA844286	Human PRO1072 (UNQ
38	726	42.8	336	22	AAU29114	Human PRO polypept
39	726	42.8	336	24	ABU71202	Human PRO1072 prot
40	726	42.8	336	24	ABU65659	Human secreted/tra
41	726	42.8	336	24	ABU65992	Novel human secret
42	726	42.8	336	24	ABU67496	Human secreted/tra
43	726	42.8	336	24	ABU61116	Human PRO1072 poly
44	726	42.8	336	24	ABU65354	Human PRO polypept
45	726	42.8	336	24	ABU58490	Human PRO polypept

ALIGNMENTS

RESULT 1	AA199372	AA199372 standard; Protein; 331 AA.
ID	AA199372;	
AC	AA199372;	
XX		
DT	08-AUG-2000 (first entry)	
XX		
DE	Human PRO1430 (UNQ736) amino acid sequence SEQ ID NO:116.	
XX		
XX	Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;	
KW	transmembrane; secretion; immunoadhesion; pharmaceutical; screening.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200012708-A2.	
XX		
PD	09-MAR-2000.	
PF	01-SNP-1999;	99WO-US20111.
XX		
XX	01-SNP-1998;	98US-0098716.
PR	01-SNP-1998;	98US-0098749.
PR	01-SNP-1998;	98US-0098750.
PR	02-SNP-1998;	98US-0098803.
PR	02-SNP-1998;	98US-0098821.
PR	02-SNP-1998;	98US-0098843.
PR	09-SNP-1998;	98US-0099536.
PR	09-SNP-1998;	98US-0099596.
PR	09-SNP-1998;	98US-0099598.
PR	09-SNP-1998;	98US-0099602.
PR	09-SNP-1998;	98US-0099642.

PR 10-SEP-1998; 98US-0099741.
 PR 10-SEP-1998; 98US-0099754.
 PR 10-SEP-1998; 98US-0099763.
 PR 10-SEP-1998; 98US-0099792.
 PR 10-SEP-1998; 98US-0099808.
 PR 10-SEP-1998; 98US-0099812.
 PR 10-SEP-1998; 98US-0099815.
 PR 10-SEP-1998; 98US-0099816.
 PR 15-SEP-1998; 98US-0100386.
 PR 15-SEP-1998; 98US-0100388.
 PR 15-SEP-1998; 98US-0100390.
 PR 16-SEP-1998; 98US-0100584.
 PR 16-SEP-1998; 98US-0100627.
 PR 16-SEP-1998; 98US-0100661.
 PR 16-SEP-1998; 98US-0100662.
 PR 16-SEP-1998; 98US-0100664.
 PR 17-SEP-1998; 98US-0100683.
 PR 17-SEP-1998; 98US-0100684.
 PR 17-SEP-1998; 98US-0100710.
 PR 17-SEP-1998; 98US-0100711.
 PR 17-SEP-1998; 98US-0100919.
 PR 17-SEP-1998; 98US-0100930.
 PR 18-SEP-1998; 98US-0100848.
 PR 18-SEP-1998; 98US-0100849.
 PR 18-SEP-1998; 98US-0101014.
 PR 18-SEP-1998; 98US-0101068.
 PR 18-SEP-1998; 98US-0101071.
 PR 22-SEP-1998; 98US-0101279.
 PR 23-SEP-1998; 98US-0101471.
 PR 23-SEP-1998; 98US-0101472.
 PR 23-SEP-1998; 98US-0101474.
 PR 23-SEP-1998; 98US-0101475.
 PR 23-SEP-1998; 98US-0101476.
 PR 23-SEP-1998; 98US-0101477.
 PR 23-SEP-1998; 98US-0101479.
 PR 24-SEP-1998; 98US-0101741.
 PR 24-SEP-1998; 98US-0101743.
 PR 24-SEP-1998; 98US-0101915.
 PR 24-SEP-1998; 98US-0101916.
 PR 29-SEP-1998; 98US-0102207.
 PR 29-SEP-1998; 98US-0102240.
 PR 29-SEP-1998; 98US-0102307.
 PR 29-SEP-1998; 98US-0102330.
 PR 29-SEP-1998; 98US-0102331.
 PR 30-SEP-1998; 98US-0102484.
 PR 30-SEP-1998; 98US-0102487.
 PR 30-SEP-1998; 98US-0102570.
 PR 30-SEP-1998; 98US-0102571.
 PR 01-OCT-1998; 98US-0102684.
 PR 01-OCT-1998; 98US-0102687.
 PR 02-OCT-1998; 98US-0102965.
 PR 06-OCT-1998; 98US-0103258.
 PR 06-OCT-1998; 98US-0103449.
 PR 07-OCT-1998; 98US-0103314.
 PR 07-OCT-1998; 98US-0103315.
 PR 07-OCT-1998; 98US-0103328.
 PR 07-OCT-1998; 98US-0103385.
 PR 07-OCT-1998; 98US-0103386.
 PR 07-OCT-1998; 98US-0103387.
 PR 07-OCT-1998; 98US-0103401.
 PR 08-OCT-1998; 98US-0103673.
 PR 08-OCT-1998; 98US-0103678.
 PR 08-OCT-1998; 98US-0103679.
 PR 08-OCT-1998; 98US-0103711.
 PR 14-OCT-1998; 98US-0104257.
 PR 20-OCT-1998; 98US-0104987.
 PR 20-OCT-1998; 98US-0105000.
 PR 20-OCT-1998; 98US-0105002.
 PR 21-OCT-1998; 98US-0105104.
 PR 22-OCT-1998; 98US-0105169.
 PR 22-OCT-1998; 98US-0105266.
 PR 26-OCT-1998; 98US-0105693.
 PR 26-OCT-1998; 98US-0105694.

PR 27-OCT-1998; 98US-0105807.
 PR 27-OCT-1998; 98US-0105881.
 PR 27-OCT-1998; 98US-0105882.
 PR 27-OCT-1998; 98US-0106062.
 PR 28-OCT-1998; 98US-0106023.
 PR 28-OCT-1998; 98US-0106029.
 PR 28-OCT-1998; 98US-0106030.
 PR 28-OCT-1998; 98US-0106032.
 PR 28-OCT-1998; 98US-0106033.
 PR 28-OCT-1998; 98US-0106178.
 PR 28-OCT-1998; 98US-0106248.
 PR 29-OCT-1998; 98US-0106386.
 PR 29-OCT-1998; 98US-0108500.
 PR 30-OCT-1998; 98US-0106464.
 PR 03-NOV-1998; 98US-0106856.
 PR 03-NOV-1998; 98US-0106902.
 PR 03-NOV-1998; 98US-0106905.
 PR 03-NOV-1998; 98US-0106919.
 PR 03-NOV-1998; 98US-0106932.
 PR 03-NOV-1998; 98US-0106934.
 PR 10-NOV-1998; 98US-0107783.
 PR 17-NOV-1998; 98US-0108775.
 PR 17-NOV-1998; 98US-0108779.
 PR 17-NOV-1998; 98US-0108787.
 PR 17-NOV-1998; 98US-0108788.
 PR 17-NOV-1998; 98US-0108801.
 PR 17-NOV-1998; 98US-0108802.
 PR 17-NOV-1998; 98US-0108806.
 PR 17-NOV-1998; 98US-0108807.
 PR 17-NOV-1998; 98US-0108867.
 PR 17-NOV-1998; 98US-0108875.
 PR 18-NOV-1998; 98US-0108848.
 PR 18-NOV-1998; 98US-0108849.
 PR 18-NOV-1998; 98US-0108850.
 PR 18-NOV-1998; 98US-0108851.
 PR 18-NOV-1998; 98US-0108852.
 PR 18-NOV-1998; 98US-0108858.
 PR 18-NOV-1998; 98US-0108904.
 (GENTECH) GENENTECH INC.
 Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
 WPI, 2000-237871/20.
 DR N-PSDB; AAA37054.
 XX New mammalian DNA sequences encoding transmembrane, receptor or
 PT secreted PRO polypeptides, useful for screening of potential peptide or
 PT small molecule inhibitors of the relevant receptor/ligand interactions
 XX Claim 12; Fig 66; 773pp; English.
 PS AAA37022 to AAA37144 encode the new isolated human transmembrane,
 XX receptor or secreted PRO polypeptides given in AA59340 to AA59462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences
 CC encoding them have various industrial applications, including uses as
 CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.
 XX
 XX Sequence 331 AA;
 SQ
 Query Match 100.0%; Score 1695; DB 21; Length 331;
 Best Local Similarity 100.0%; Pred. No. 4.5e-166;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSRYLLPLSALGVAGAVILIKDYVTGACPSKATIPKTVIVTANTGIGQTALELAR 60
 Db 1 MSRYLLPLSALGVAGAVILIKDYVTGACPSKATIPKTVIVTANTGIGQTALELAR 60
 QY 61 RGNITILACDMECEAAADIRGETLNHHVNAHRLDLASLSIREFAKITEEBRVDI 120

CC be used for treating disorders of colon, brain, skin, heart, blood
 CC vessels, kidney, prostate, skeletal muscle, ovary, testis and epididymis,
 CC spleen, lung, liver, uterus and endometrium, T-cells, thymus,
 CC B cells, breast, thyroid, pancreas, small intestine, red blood platelet
 CC number, precursor T cell neoplasms, bone forming cells, and bone marrow
 CC cells.

XX Sequence 331 AA;

Query Match 99.8%; Score 1691; DB 22; Length 331;
 Best Local Similarity 99.7%; Pred. No. 1.2e-165;

Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSRYLLPLSALGTAGAAVLLKDYVTGACPSKATIPGKTYIVTGANTGIGKOTALELAR 60
 DB 1 MSRYLLPLSALGTAGAAVLLKDYVTGACPSKATIPGKTYIVTGANTGIGKOTALELAR 60
 QY 61 RGNITILACRDMKCEAAADIRGETLNHHVNARHDLASIKSIREPAKIIIEBERVDI 120
 DB 61 RGNITILACRDMKCEAAADIRGETLNHHVNARHDLASIKSIREPAKIIIEBERVDI 120
 QY 121 LINNAGVWMCPTHITTEDEGFMQGVNHLGHFLTNLLDLKXKASAPSRRIINLSLAHYAG 180
 DB 121 LINNAGVWMCPTHITTEDEGFMQGVNHLGHFLTNLLDLKXKASAPSRRIINLSLAHYAG 180
 QY 181 HIDPDLNMQTRKYNKKAACOSKLAIVLFTKELSRLOQSGGVYNALHPGARTELGRH 240
 DB 181 HIDPDLNMQTRKYNKKAACOSKLAIVLFTKELSRLOQSGGVYNALHPGARTELGRH 240
 QY 241 TGIHSTFSSSTTIGPIFWLIVKSPPELAAQSTYLAVAEELADVGKTYFDGLKOKAPAPEA 300
 DB 241 TGIHSTFSSSTTIGPIFWLIVKSPPELAAQSTYLAVAEELADVGKTYFDGLKOKAPAPEA 300
 QY 301 EDEEVARRIMASARLVGLEAPSVREQLPR 331
 DB 301 EDEEVARRIMASARLVGLEAPSVREQLPR 331

4 RESULT 5
 AAU76222 standard; Protein; 331 AA.

XX AAU76222;

XX 08-MAY-2002 (first entry)

DE Human 21676 alcohol dehydrogenase (ADH) protein.

XX Alcohol dehydrogenase; ADH; human; cytosolic; antiinflammatory;
 KW cerebrioprotective; anti-HIV; immunomodulator; hepatotropic; metastases;
 KW pulmonary congestion; Meckel diverticulum; splenic infarction;
 KW idiopathic inflammatory bowel disease; jaundice; cholestasis;
 KW endometriosis; cerebral edema; AIDS; leukopenia; splenomegaly;
 KW acquired immune deficiency disease; lupus erythematosus; dermatitis;
 KW lung disease; adult respiratory distress syndrome; skin disease;
 KW bronchitis; sarcoidosis; pneumothorax; colon disorder; colitis;
 KW Crohn's disease; liver disorder; hepatitis; cirrhosis; brain disorder;
 KW meningitis; Alzheimer's disease; Huntington's disease; atherosclerosis;
 KW ischaemia; 21676; enzyme.

XX Homo sapiens.

OS US2002010946-A1.

XX 24-JAN-2002.

XX 28-FEB-2001; 2001US-0796089.

XX 15-DEC-1999; 99US-0464039.

XX 15-DEC-2000; 2000WO-US33873.

XX (MILL-) MILLENNIUM PHARM INC.

PI Meyers R;
 XX WPI; 2002-179233/23.
 DR N-PSDB; ABR15714.

XX New human alcohol dehydrogenase (ADH) polynucleotides and polypeptides,
 PT useful as targets for diagnosing or treating ADH-related or
 PT ADH-mediated disorders, e.g. malignant breast metastases, anemia or
 PT leukopenia

XX Claim 9; Fig 11A-11B; 80pp; English.

XX This invention relates to the cDNA and protein sequences of 5 novel
 CC human alcohol dehydrogenase molecules. The ADH polynucleotides and
 CC polypeptides are useful as targets for diagnosing or treating ADH-
 CC related or ADH-mediated disorders, e.g. malignant breast, liver, colon
 CC or liver metastases, pulmonary congestion or edema, Meckel diverticulum,
 CC idiopathic inflammatory bowel disease, jaundice and cholestasis,
 CC endometriosis, cerebral edema, AIDS, or leukopenia. The sequences
 CC may also be used for treating other diseases or disorders such as spleen
 CC disorders (splenomegaly, splenic infarction), lung diseases (adult
 CC respiratory distress syndrome, bronchitis, sarcoidosis, pneumothorax),
 CC colon disorders (colitis, Crohn's disease), liver disorders (hepatitis,
 CC cirrhosis), brain disorders (meningitis, Alzheimer's disease,
 CC Huntington's disease), heart and blood vessel disorders
 CC (atherosclerosis, ischaemia), skin diseases (lupus erythematosus,
 CC dermatitis) and many other diseases listed in the specification. The
 CC polynucleotides and polypeptides are also useful in screening methods to
 CC identify agonists and antagonists for diagnosis or treatment. In
 CC particular, the polypeptides and polynucleotides are useful in drug
 CC screening assays in cell-based assays or cell-free systems, as well as
 CC for biological assays related to ADHs. The ADH polypeptides are also
 CC useful for producing antibodies specific for the ADH regions. The
 CC polynucleotides and polypeptides may also be used for monitoring
 CC therapeutic effects during clinical trials and other treatments.
 CC The present sequence represents the human 21676 alcohol dehydrogenase
 CC protein of the invention.

XX Sequence 331 AA;

Query Match 99.8%; Score 1691; DB 23; Length 331;
 Best Local Similarity 99.7%; Pred. No. 1.2e-165;
 Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSRYLLPLSALGTAGAAVLLKDYVTGACPSKATIPGKTYIVTGANTGIGKOTALELAR 60
 DB 1 MSRYLLPLSALGTAGAAVLLKDYVTGACPSKATIPGKTYIVTGANTGIGKOTALELAR 60
 QY 61 RGNITILACRDMKCEAAADIRGETLNHHVNARHDLASIKSIREPAKIIIEBERVDI 120
 DB 61 RGNITILACRDMKCEAAADIRGETLNHHVNARHDLASIKSIREPAKIIIEBERVDI 120
 QY 121 LINNAGVWMCPTHITTEDEGFMQGVNHLGHFLTNLLDLKXKASAPSRRIINLSLAHYAG 180
 DB 121 LINNAGVWMCPTHITTEDEGFMQGVNHLGHFLTNLLDLKXKASAPSRRIINLSLAHYAG 180
 QY 181 HIDPDLNMQTRKYNKKAACOSKLAIVLFTKELSRLOQSGGVYNALHPGARTELGRH 240
 DB 181 HIDPDLNMQTRKYNKKAACOSKLAIVLFTKELSRLOQSGGVYNALHPGARTELGRH 240
 QY 241 TGIHSTFSSSTTIGPIFWLIVKSPPELAAQSTYLAVAEELADVGKTYFDGLKOKAPAPEA 300
 DB 241 TGIHSTFSSSTTIGPIFWLIVKSPPELAAQSTYLAVAEELADVGKTYFDGLKOKAPAPEA 300
 QY 301 EDEEVARRIMASARLVGLEAPSVREQLPR 331
 DB 301 EDEEVARRIMASARLVGLEAPSVREQLPR 331

RESULT 6
 AAB88356 standard; Protein; 331 AA.

XX AAB88356

AC	AAB88356;
XX	
DT	23-MAY-2001 (first entry)
DE	Human membrane or secretory protein clone PSEC0082.
XX	
XX	Human; secretory protein; membrane protein; vaccine; gene therapy;
KW	rheumatoid arthritis; diabetes.
OS	Homo sapiens.
XX	
PN	EPI067182-A2.
XX	
PD	10-JAN-2001.
PF	07-JUL-2000; 2000EP-0114090.
XX	
PR	08-JUL-1999; 99JP-0194179.
PR	11-JAN-2000; 2000JP-0118775.
PR	02-MAY-2000; 2000JP-0183766.
PA	(HELI-) HELIX RES INST.
XX	
PI	Oca T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
DR	WPI; 2001-093989/11.
DR	N-PSDB; AAF93783.
XX	
PT	Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT	gene therapy or as candidate target molecules in drug development -
PS	Claim 1; SEQ ID 80; 609bp + CD ROM; English.
XX	
XX	This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC	which encode human secretory or membrane proteins represented by
CC	AAB88317 - AAB88419. Included in the invention are primers
CC	AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC	cDNA sequences of the invention. The invention also includes methods for
CC	the production of antibodies directed against the proteins, and cDNA
CC	sequences, which can be used in vaccines. The polynucleotide sequences
CC	can be used in gene therapy. The polynucleotide sequences and the
CC	proteins they encode may be used in the prevention, treatment and
CC	diagnosis of diseases associated with inappropriate secretory
CC	protein/membrane protein expression. The nucleic acids and complementary
CC	sequences may also be used as DNA probes in diagnostic assays
CC	(e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC	presence of similar nucleic acid sequences in samples. They may also be
CC	used to study the expression and function of secretory proteins/membrane
CC	polypeptides and their role in metabolism. The polypeptides may be used
CC	as antigens in the production of antibodies against them and in assays to
CC	identify modulators (agonists and antagonists) of expression and
CC	activity. The antibodies and antagonists may also be used as therapeutic
CC	agents to down regulate expression and activity. The antibodies may also
CC	be used as diagnostic agents for detecting the presence of the
CC	polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC	[ELISA]. Examples of diseases which may be treated include rheumatoid
CC	arthritis and diabetes.
XX	
SQ	Sequence 331 AA;
QY	Query Match 99.6%; Score 1688; DB 22; Length 331;
Bd	Best Local Similarity 99.7%; Pred. No. 2,4e-165;
Md	Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 MSRYLLPISALGTGVAATLKDYYTGAGCPSPKATIPGVITVGANTGIGXOTALELAR 60
Dd	1 MSRYLLPISALGTGVAATLKDYYTGAGCPSPKATIPGVITVGANTGIGXOTALELAR 60
QY	61 RCGNIIILACRDMKECEAAAKDIRGETLNHNVNARHDLASIKSIREFPAKIIEEERVDI 120
Dd	61 RCGNIIILACRDMKECEAAAKDIRGETLNHNVNARHDLASIKSIRVPAKIIIEEERVDI 120
QY	121 LNNAGVMCEPHMTTDEGEFMGVNHGHFLITNLILDLYKASAPSRITINLSLAHVAG 180

Accession	Protein Name	Location/Qualifiers
Dd	121 LINAGVRCRHWITTEDEFGVGNHGHFLITLLDLKLAAAPSRIINISLAHVAG	180
Qy	181 HIDEEDLNMQTRKNTKAAVCGSKLAIYLFTEKLSRRLOGGGVYVNAHPGVAFTELGRH	240
Dd	181 HIDFDLNMQRKNTKAAVCGSKLAIYLFTEKLSRRLOGGGVYVNAHPGVAFTELGRH	240
Qy	241 TGIHGSTSSITLLGPIFWLLVKSPELAQPSITYLAVAEELAVSGVCPFGLLKOKAPAPAA	300
Dd	241 TGIHGSTSSITLLGPIFWLLVKSPELAQPSITYLAVAEELAVSGVCPFGLLKOKAPAPAA	300
Qy	301 EDEEVARRLMAESARLVGLEAPSVAEOPLEPR	331
Dd	301 EDEEVARRLMAESARLVGLEAPSVAEOPLEPR	331
RESULT 7		
ID	AAAB4365 standard; Protein; 260 AA.	
AC	AAAB4365;	
DT	22-AUG-2001 (first entry)	
DE	Amino acid sequence of human alcohol dehydrogenase 33756.	
XX	Human; alcohol dehydrogenase; colon disorder; brain disorder;	
XX	skin disorder; heart disorder; blood vessel disorder; kidney disorder;	
XX	prostate disorder; skeletal muscle disorder; ovary disorder;	
XX	testis disorder; epididymis disorder; spleen disorder; lung disorder;	
XX	liver disorder; uterus disorder; endometrium disorder; T-cell disorder;	
XX	red cell disorder; thymus disorder; B cell disorder; breast disorder;	
XX	thyroid disorder; pancreas disorder; small intestine disorder;	
XX	reduced platelet number disorder; precursor T cell neoplasm.	
OS	Homo sapiens.	
XX	Key	
FM	Modified-site	Location/Qualifiers
FM	Modified-site	29..31
FM	Modified-site	/note= "putative protein kinase C phosphorylation site"
FM	Modified-site	32..34
FM	Modified-site	/note= "putative protein kinase C phosphorylation site"
FM	Modified-site	32..35
FM	Modified-site	/note= "putative casein kinase II phosphorylation site"
FM	Modified-site	61..66
FM	Modified-site	/note= "putative casein kinase II phosphorylation site"
FM	Modified-site	100..103
FM	Modified-site	/note= "putative N-glycosylation site"
FM	Modified-site	120..122
FM	Modified-site	/note= "putative protein kinase C phosphorylation site"
FM	Modified-site	144..146
FM	Modified-site	/note= "putative protein kinase C phosphorylation site"
FM	Modified-site	149..154
FM	Modified-site	/note= "putative N-myristoylation site"
FM	Modified-site	160..165
FM	Modified-site	/note= "putative N-myristoylation site"
FM	Modified-site	171..176
FM	Modified-site	/note= "putative N-myristoylation site"
FM	Modified-site	213..215
FM	Modified-site	/note= "putative protein kinase C phosphorylation site"
FM	Modified-site	232..244
FM	Modified-site	/note= "putative protein kinase C phosphorylation site"
FM	Modified-site	251..254
FM	Modified-site	/note= "putative protein kinase C phosphorylation site"
FM	Modified-site	251..255
FM	Modified-site	/note= "putative casein kinase II phosphorylation site"
XX	MO200144446-A2.	
XX	21-JUN-2001.	
XX	15-DEC-2000. 2000MO-US33873.	

PR 15-DEC-1999; 99US-0464039.
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX
XX
PI Meyers R;
XX WPI; 2001-390244/41.
DR N-PSDB; AAK25132.
XX
PT Novel human alcohol dehydrogenase proteins, 21612, 21615, 21620, 21676,
PT 33756, useful for treating psoriasis, tropical sprue, pancreatitis,
PT goiter, osteomalacia, endometriosiis, angina pectoris, embolism
XX
XX
PS Claim 9; Fig 7; 156pp; English.
XX
XX
CC AAB8464-68 represent human alcohol dehydrogenase proteins, designated
CC 21620, 33756, 21676, 21612 and 21615, respectively. Alcohol dehydrogenase
CC polynucleotides and polypeptides are useful for treatment and diagnosis
CC of disorders mediated by or related to alcohol dehydrogenase. They can
CC be used for treating disorders of colon, brain, skin, heart, blood
CC vessels, kidney, prostate, skeletal muscle, ovary, testis and epididymis,
CC spleen, lung, liver, uterus and endometrium, T-cells, red cells, thymus,
CC B cells, breast, thyroid, pancreas, small intestine, reduced platelet
CC number, precursor T cell neoplasms, bone forming cells, and bone marrow
CC cells.
XX
XX
SQ Sequence 260 AA;
Query Match 79.1%; Score 1340; DB 22; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.4e-129;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 72 MEKCEAAKDIRGETLNHNHVARHLDLASKSIREFAKIIIEBERVDILINNAGVWRCP 131
DB 1 MEKCEAAKDIRGETLNHNHVARHLDLASKSIREFAKIIIEBERVDILINNAGVWRCP 60
QY 132 HMTTEDEGFEMQFVNHLGHFLITNLNLDKTKASAPSRITNLSLAHVAGHIDPDDLNMOT 191
DB 61 HMTTEDEGFEMQFVNHLGHFLITNLNLDKTKASAPSRITNLSLAHVAGHIDPDDLNMOT 120
QY 192 RKVYTKAYCOSKAIIVFTKELSRRLQSGGVTVNAALHPGVARTELGRHTGHSFSSST 251
DB 121 RKVYTKAYCOSKAIIVFTKELSRRLQSGGVTVNAALHPGVARTELGRHTGHSFSSST 180
QY 252 TLGPIFMLIVSPELAAQPSSTYLAVAEELADVSGKYFDGLKKAPAPAEDEEVARLMA 311
DB 181 TLGPIFMLIVSPELAAQPSSTYLAVAEELADVSGKYFDGLKKAPAPAEDEEVARLMA 240
QY 312 ESARLVGLEAPSVREQPLPR 331
DB 241 ESARLVGLEAPSVREQPLPR 260
RESULT 8
AAU76221
ID AAU76221 standard; Protein; 260 AA.
XX
AC AAU76221;
XX
XX
DT 08-MAY-2002 (first entry)
XX
XX Human 33756 alcohol dehydrogenase (ADH) protein.
KM Alcohol dehydrogenase; ADH; human; cytosolic; antiinflammatory;
KM cerebroprotective; anti-HIV; immunomodulator; hepatocytotoxic; metatases;
KM pulmonary congestion; Meckel diverticulum; splenic infection;
KM idiopathic inflammatory bowel disease; jaundice; cholestasis;
KM endometriosis; cerebral edema; AIDS; leukopenia; splenomegaly;
KM acquired immune deficiency disease; lupus erythematosus; dermatitis;
KM lung disease; adult respiratory distress syndrome; skin disease;
KM bronchitis; sarcoidosis; pneumothorax; colon disorder; colitis;
KM Crohn's disease; liver disorder; hepatitis; cirrhosis; brain disorder;
KM meningitis; Alzheimer's disease; Huntington's disease; atherosclerosis;

KM ischaemia; 33756; enzyme.
XX
XX
OS Homo sapiens.
XX
XX
PN US2002010946-A1.
XX
XX
PD 24-JAN-2002.
XX
XX
PF 28-FEB-2001; 2001US-0796089.
XX
XX
PR 15-DEC-1999; 99US-0464039.
PR 15-DEC-2000; 2000WO-US33873.
XX
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
XX
PI Meyers R;
XX
XX
DR WPI; 2002-179233/23.
DR N-PSDB; ABK15713.
XX
XX
PT New human alcohol dehydrogenase (ADH) polynucleotides and polypeptides,
PT useful as targets for diagnosing or treating ADH-related or
PT ADH-mediated disorders, e.g. malignant breast metastases, enema or
PT leukopenia
XX
XX
PS Claim 9; Fig 7; 80pp; English.
XX
XX
CC This invention relates to the cDNA and protein sequences of 5 novel
CC human alcohol dehydrogenase molecules. The ADH polynucleotides and
CC polypeptides are useful as targets for diagnosing or treating ADH-
CC related or ADH-mediated disorders, e.g. malignant breast, liver, colon
CC or liver metastases, pulmonary congestion or enema, Meckel diverticulum,
CC idiopathic inflammatory bowel disease, jaundice and cholestasis,
CC endometriosis, cerebral edema, AIDS, or leukopenia. The sequences
CC may also be used for treating other diseases or disorders such as spleen
CC disorders (splenomegaly, splenic infarction), lung diseases (adult
CC respiratory distress syndrome, bronchitis, sarcoidosis, pneumothorax),
CC colon disorders (colitis, Crohn's disease), liver disorders (hepatitis,
CC cirrhosis), brain disorders (meningitis, Alzheimer's disease,
CC Huntington's disease), heart and blood vessel disorders
CC (atherosclerosis, ischaemia), skin diseases (lupus erythematosus,
CC dermatitis) and many other diseases listed in the specification. The
CC polynucleotides and polypeptides are also useful in screening methods to
CC identify agonists and antagonists for diagnosis or treatment. In
CC particular, the polypeptides and polynucleotides are useful in drug
CC screening assays in cell-based assays or cell-free systems, as well as
CC for biological assays related to ADHs. The ADH polypeptides are also
CC useful for producing antibodies specific for the ADH regions. The
CC polynucleotides and polypeptides may also be used for monitoring
CC therapeutic effects during clinical trials and other treatments.
CC The present sequence represents the human 21620 alcohol dehydrogenase
CC protein of the invention.
XX
XX
SQ Sequence 260 AA;
Query Match 79.1%; Score 1340; DB 23; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.4e-129;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 72 MEKCEAAKDIRGETLNHNHVARHLDLASKSIREFAKIIIEBERVDILINNAGVWRCP 131
DB 1 MEKCEAAKDIRGETLNHNHVARHLDLASKSIREFAKIIIEBERVDILINNAGVWRCP 60
QY 132 HMTTEDEGFEMQFVNHLGHFLITNLNLDKTKASAPSRITNLSLAHVAGHIDPDDLNMOT 191
DB 61 HMTTEDEGFEMQFVNHLGHFLITNLNLDKTKASAPSRITNLSLAHVAGHIDPDDLNMOT 120
QY 192 RKVYTKAYCOSKAIIVFTKELSRRLQSGGVTVNAALHPGVARTELGRHTGHSFSSST 251
DB 121 RKVYTKAYCOSKAIIVFTKELSRRLQSGGVTVNAALHPGVARTELGRHTGHSFSSST 180
QY 252 TLGPIFMLIVSPELAAQPSSTYLAVAEELADVSGKYFDGLKKAPAPAEDEEVARLMA 311

Db	181	ILGPIFWLLVSPSELLAQSTYTLAAVAEELADVSGKRFQGLKKAPAPAEADEVARRLMA	244
Cy	312	ESARLVGLEAPSVREQPLPR	331
Db	241	ESARLVGLEAPSVREQPLPR	260
RESULT 9			
ID	AAW40387		
AC	AAW40387 standard; Protein; 260 AA.		
XX	AAW40387;		
XX			
DT	22-OCT-2001 (first entry)		
DE	Human polypeptide SEQ ID NO 3532.		
KM	Human; nocropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KM	peripheral nervous system; neuropathy; central nervous system; CNS;		
KM	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KM	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KM	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KM	leukaemia.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000WO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
XX	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HSE-) HSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RF;		
XX			
DR	WPI; 2001-442253/47.		
DR	N-PSDB; AA159543.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
XX	such as central nervous system injuries -		
XX			
XX	Example 7; SEQ ID NO 3532; 1007bp; English.		
XX			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AAW38642-AAW42213) with nocropic,		
CC	immunosuppressant and cyostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX			
XX	Sequence 260 AA;		
XX			

	Query Match	78.6%	Score 1332	DB 22	Length 260	
	Best Local Similarity	99.6%	Pred. No. 9.3e-129			
	Matches 259	Conservative	0	Mismatches 1	Indels 0	Gaps 0
Qy	72 MEKCEAAAKDTRGETLNNHVNARHLDIASLSIKIREFAKTIIEBERVDILINAGVRCF	131				
Dd	1 MEKCEAAAKDTRGETLNNHVNARHLDIASLSIKIREFAKTIIEBERVDILINAGVRCF	60				
Qy	132 HWTEDGFENQPGYNHLGFFLLTNLLIDKTKASPSSIIINLSLAHYAGHIDFDPLWQT	191				
Dd	61 HWTEDESEMQFGVNHGHFFLLTNLLIDKTKASPSSIIINLSLAHYAGHIDFDPLWQT	120				
Qy	192 RKNTTKAAYCOSKIALYLFKEISLRQSGSVYNAHPGVARELGRHGTHGSTSST	251				
Dd	121 RKNTTKAAYCOSKIALYLFKEISRRLQSGSVYNAHPGVARELGRHGTHGSTSST	180				
Qy	252 TLGITFPLLVSPELTAQPSTYLAVAEELADVSGKYFDGLKKQKPAPAEDEEVARRLMA	311				
Dd	181 TLGITFPLLVSPELTAQPSTYLAVAEELADVSGKYFDGLKKQKPAPAEDEEVARRLMA	240				
Qy	312 ESARLVGLEAPSVEQEPLEPR 331					
Dd	241 ESARLVGLEAPSVEQEPLEPR 260					
	RESULT 10					
ID	AAE03299					
XX	AAE03299 standard; Protein: 240 AA.					
AC	AAE03299;					
XX						
DT	10-AUG-2001 (first entry)					
XX						
XX	Human gene 8 encoded secreted protein HPCK10, SEQ ID NO:73.					
Dd						
KM	Human; secreted protein; proliferative disorder; cancer; tumour;					
KM	fœtal abnormality; developmental abnormality; haematopoietic disorder;					
KM	immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;					
KM	inflammation; allergy; neurological disorder; Alzheimer's disease;					
KM	Parinson's disease; cognitive disorder; schizophrenia; asthma;					
KM	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;					
KM	cardiovascular disorder; angiotenic disorder; kidney disorder;					
KM	gastrointestinal disorder; pregnancy-related disorder;					
KM	endocrine disorder; infection; wound healing; vulnery;					
KM	cell culture; chemotaxis; food additive; gene therapy;					
KM	binding partner identification.					
OS	Homo sapiens.					
XX						
FH	Key	Location/Qualifiers				
FT	Peptide	1..16				
FT	Protein	/label= Signal_peptide 17..240				
FT		/note= "Mature secreted protein"				
XX						
PN	WO200134800-A1.					
PD	17-MAY-2001.					
XX						
PF	08-NOV-2000; 2000MO-US30674.					
XX						
PR	12-NOV-1999; 99US-0164750.					
PR	30-JUN-2000; 2000US-025128.					
XX						
PA	(HDWA-) HUMAN GENOME SCI INC.					
XX						
PI	Ruben SM, Komatsoulis GA, Ebner R, Fiacella M, Wei P;					
DR	WPI, 2001-329085/34.					
DR	N-PsDB; AAD07712.					
XX						
PT	New nucleic acid molecules encoding human secreted proteins, used in					

PT Preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

PS Claim 11; Page 461-462; 530pp; English.

CC AAD07705-AAD07759 represent cDNAs corresponding to 19 human secreted protein genes, and AAE03292-AAE03346 represent the proteins they encode. CC AAE03347-AAE03375 represent human secreted protein fragments or variants. CC The genes and their secreted proteins are useful for preventing, CC treating or ameliorating medical conditions, e.g., by protein or gene therapy. CC Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. CC Specific uses are described for each of the 19 genes, based on the tissues in which they are most highly expressed, CC and include developing products for the diagnosis or treatment of CC proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, CC allergies, neurological disorders (e.g., Alzheimer's disease, CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, CC cardiovascular disorders, angiogenic disorders, kidney disorders, CC gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. CC The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used CC as a food additive or preservative to modify storage properties. CC Antibodies specific for a protein of the invention can be used in CC alleviating symptoms associated with the disorders mentioned above, and CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein of the invention.

XX Sequence 240 AA;

Query Match 67.0%; Score 1136; DB 22; Length 240;

Best Local Similarity 97.8%; Pred. No. 1.4e-108; Matches 22; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSRYLPALSGTGAAGVLLKDYVTGACPSKATIPGKTVVTGANTGIGKGTALFLAR 60

DB 1 MSRYLPALSGTGAAGVLLKDYVTGACPSKATIPGKTVVTGANTGIGKGTALFLAR 60

QY 61 RGNIIILACRDMCEAAANDIRGETLNHHVNAHLDIASLSIREFAKIIIEEBRVDI 120

DB 61 RGNIIILACRDMCEAAANDIRGETLNHHVNAHLDIASLSIREFAKIIIEEBRVDI 120

QY 121 LINNAGVRCPHMTTDEGFEWQGVNHLGHFLINLLDKAKASPSRIINLSLAHVAG 180

DB 121 LINNAGVRCPHMTTDEGFEWQGVNHLGHFLINLLDKAKASPSRIINLSLAHVAG 180

QY 181 HIDPDDINMQRKNTKRAAYQSKLAIYLFKELSRRIOGSGVTYNA 227

DB 181 HIDPDDINMQRKNTKRAAYQSKLAIYLFKELSRRIOGSGVTYNA 227

RESULT 11

AAE01684

ID AAE01684 standard; Protein; 240 AA.

AC AAE01684;

XX 18-JUL-2001 (first entry)

DE Human gene 13 encoded secreted protein HLEDB16, SEQ ID NO:96.

XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma; foetal abnormality; developmental abnormality; haematopoietic disorder; immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis; Parkinson's disease; cognitive disorder; schizophrenia; skin disorder; psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;

KW inflammation; neurological disorder; Alzheimer's disease; food additive; KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy; KW pregnancy-related disorder; endocrine disorder; infection; wound healing; KW cell culture; chemotaxis; vulnery; binding partner identification; KW gene therapy.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..16

FT Protein /label= signal_peptide

XX /label= Mature_human_secreted_protein

PD 17-MAY-2001.

XX 01-NOV-2000; 2000WO-US30036.

XX 05-NOV-1999; 99US-0163576.

XX 27-JUL-2000; 2000US-0221366.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Soppet DR, Komatsoulis G, Shi Y, Olsen HS, Ruben SM;

DR WPI, 2001-316492/33.

DR N-PSDB; AAD05504.

PT Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -

XX Claim 11; Page 475-476; 540pp; English.

XX AAD05492-AAD05564 represent cDNAs corresponding to 22 human secreted protein genes, and AAE01672-AAE01743 represent the proteins they encode. CC AAE01744-AAE01763 represent human secreted protein fragments or variants. CC The secreted proteins and their genes are useful for preventing, treating CC or ameliorating medical conditions, e.g., by protein or gene therapy. CC Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. CC Specific uses are described for each of the 22 genes, CC based on the tissues in which they are most highly expressed, and include CC developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, CC angiogenic disorders, kidney disorders, gastrointestinal disorders, CC pregnancy-related disorders, endocrine disorders, and infections. CC The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs CC before transplantation, for supporting cell culture of primary tissues, CC to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or CC preservative to modify storage properties. Antibodies specific for a CC protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). CC The present sequence represents a human secreted protein of the invention.

XX Sequence 240 AA;

Query Match 67.0%; Score 1136; DB 22; Length 240;

Best Local Similarity 97.8%; Pred. No. 1.4e-108; Matches 22; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSRYLPALSGTGAAGVLLKDYVTGACPSKATIPGKTVVTGANTGIGKGTALFLAR 60

DQ 1 MSRYLLPISALGTGAGAAVLKDYTGACPSKATIPGKTVITGANTGIGKQTALIELAR 60
 QY 61 RGNIIIIACRMEKEAAAKDIRGETLNHNHVARHIDLASLSIREPAKIIIEEERVDI 120
 DB 61 RGNIIIIACRMEKEAAAKDIRGETLNHNHVARHIDLASLSIREPAKIIIEEERVDI 120
 QY 121 LINNAGVRCRPHWTTEDEGEMQPGVNHLSGHPFLTNLLDKLKASAPRIINLSLAHVAG 180
 DB 121 LINNAGVRCRPHWTTEDEGEMQPGVNHLSGHPFLTNLLDKLKASAPRIINLSLAHVAG 180
 QY 181 HIDEFDLWQTRKNTKAAVCOSKLAIVLFTKELSRRLQSGVTANA 227
 DB 181 HIDEFDLWQTRKNTKAAVCOSKLAIVLFTKELSRRLQSGVTANA 227
 RESULT 12
 ID ABG63953 standard; Protein; 240 AA.
 AC ABG63953;
 DT 27-AUG-2002 (first entry)
 DE Human albumin fusion protein #628.
 XX
 KW Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytoskeletal; antiinfectility; antiinflammatory; antitumor;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.
 OS Homo sapiens.
 OS Synthetic.
 PN WO200177137-A1.
 PD 18-OCT-2001.
 PF 12-APR-2001; 2001WO-US11988.
 PR 12-APR-2000; 2000US-229358P.
 PR 25-APR-2000; 2000US-199384P.
 PR 21-DEC-2000; 2000US-256931P.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Haseltine WA;
 DR WPI; 2002-010886/01.
 PT New fusion protein for treating disease e.g. diabetes comprises an
 PT albumin fused to a therapeutic protein -
 XX Claim 1, Page 925; 2102pp; English.
 XX The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA, also known as human serum
 CC albumin, HSA). The proteins are useful for treating a disease or
 CC disorder that may be modulated by therapeutic protein X. The albumin
 CC extends the shelf-life of protein X, and may increase its biological
 CC in vitro/in vivo activity. The protein is useful for treating and
 CC diagnosing disorders such as cancer, reproductive disorders, digestive
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 CC (e.g. diabetes), haematopoietic disorders, neural disorders
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 CC encephalomyelitis, meningitis), schizophrenia), and connective disorders
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
 CC fusion proteins of the invention.
 XX

SQ Sequence 240 AA;
 Query Match 67.0%; Score 1136; DB 23; Length 240;
 Best Local Similarity 97.8%; Pred. No. 1,4e-108;
 Matches 222; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSRYLLPISALGTGAGAAVLKDYTGACPSKATIPGKTVITGANTGIGKQTALIELAR 60
 DB 1 MSRYLLPISALGTGAGAAVLKDYTGACPSKATIPGKTVITGANTGIGKQTALIELAR 60
 QY 61 RGNIIIIACRMEKEAAAKDIRGETLNHNHVARHIDLASLSIREPAKIIIEEERVDI 120
 DB 61 RGNIIIIACRMEKEAAAKDIRGETLNHNHVARHIDLASLSIREPAKIIIEEERVDI 120
 QY 121 LINNAGVRCRPHWTTEDEGEMQPGVNHLSGHPFLTNLLDKLKASAPRIINLSLAHVAG 180
 DB 121 LINNAGVRCRPHWTTEDEGEMQPGVNHLSGHPFLTNLLDKLKASAPRIINLSLAHVAG 180
 QY 181 HIDEFDLWQTRKNTKAAVCOSKLAIVLFTKELSRRLQSGVTANA 227
 DB 181 HIDEFDLWQTRKNTKAAVCOSKLAIVLFTKELSRRLQSGVTANA 227
 RESULT 13
 ID ABG64446 standard; Protein; 240 AA.
 AC ABG64446;
 DT 27-AUG-2002 (first entry)
 DE Human albumin fusion protein #1121.
 XX
 KW Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytoskeletal; antiinfectility; antiinflammatory; antitumor;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.
 OS Homo sapiens.
 OS Synthetic.
 PN WO200177137-A1.
 PD 18-OCT-2001.
 PF 12-APR-2001; 2001WO-US11988.
 PR 12-APR-2000; 2000US-229358P.
 PR 25-APR-2000; 2000US-199384P.
 PR 21-DEC-2000; 2000US-256931P.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Haseltine WA;
 DR WPI; 2002-010886/01.
 PT New fusion protein for treating disease e.g. diabetes comprises an
 PT albumin fused to a therapeutic protein -
 XX Claim 1, Page 1218-1219; 2102pp; English.
 XX The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA, also known as human serum
 CC albumin, HSA). The proteins are useful for treating a disease or
 CC disorder that may be modulated by therapeutic protein X. The albumin
 CC extends the shelf-life of protein X, and may increase its biological
 CC in vitro/in vivo activity. The protein is useful for treating and
 CC diagnosing disorders such as cancer, reproductive disorders, digestive
 CC

CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 CC (e.g. diabetes), haematopoietic disorders, neural disorders
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
 CC (e.g. osteoporosis, arthritis). AAB63326-ABG65518 represent albumin
 CC fusion proteins of the invention.

XX
 SQ Sequence 240 AA;

Query Match 67.0%; Score 1136; DB 23; Length 240;
 Best Local Similarity 97.8%; Pred. No. 1.4e-108;
 Matches 222; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSRYLLPLSALGTVAAGAVLLKDYTGACPSKATIPKTYIVTGANTGIGQTALELAR 60
 DB 1 MSRYLLPLSALGTVAAGAVLLKDYTGACPSKATIPKTYIVTGANTGIGQTALELAR 60
 QY 61 RGNITILACRDMKCEAAADIRGETLNHHVNAHDLASLSKSTREPAKTIIEBERVDI 120
 DB 61 RGNITILACRDMKCEAAADIRGETLNHHVNAHDLASLSKSTREPAKTIIEBERVDI 120
 QY 121 LINNAGVRCPHMTTEDEGFEMQFGVNLGHFLTLNLLDCLKASAPSRINISSLAHVAG 180
 DB 121 LINNAGVRCPHMTTEDEGFEMQFGVNLGHFLTLNLLDCLKASAPSRINISSLAHVAG 180
 QY 181 HIDFDDLNMQTRKNTKAAVCQSKLAIVLFKELSRRIQSGGVTVNA 227
 DB 181 HIDFDDLNMQTRKNTKAAVCQSKLAIVLFKELSRRIQSGGVTVNA 227

RESULT 14
 AAB57079 ID AAB57079 standard; Protein; 273 AA.
 AC AAB57079;
 XX 13-MAR-2001 (first entry)
 DT
 DE Human prostate cancer antigen protein sequence SEQ ID NO:1657.
 XX
 KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytoskeletal; cardioprotective; immunomodulatory; muscular;
 KW vulnerrary; gastrointestinal; nephrotoxic; anti-infective; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease.

XX
 OS Homo sapiens.
 XX
 PN WO200055174-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05988.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR MPI, 2000-587513/55.
 DR N-PSDB; AAF16282.
 XX
 PT Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX
 PS Claim 11; Page 2115-2116; 2338pp; English.
 XX
 CC AAF15566 to AAF16505 encode the human prostate cancer associated

CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytoskeletal,
 CC cardioprotective, immunomodulatory, muscular, vulnerary, gastrointestinal,
 CC nephrotoxic, anti-infective, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention.

XX
 SQ Sequence 273 AA;

Query Match 67.0%; Score 1136; DB 21; Length 273;
 Best Local Similarity 97.8%; Pred. No. 1.7e-108;
 Matches 222; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSRYLLPLSALGTVAAGAVLLKDYTGACPSKATIPKTYIVTGANTGIGQTALELAR 60
 DB 34 MSRYLLPLSALGTVAAGAVLLKDYTGACPSKATIPKTYIVTGANTGIGQTALELAR 93
 QY 61 RGNITILACRDMKCEAAADIRGETLNHHVNAHDLASLSKSTREPAKTIIEBERVDI 120
 DB 94 RGNITILACRDMKCEAAADIRGETLNHHVNAHDLASLSKSTREPAKTIIEBERVDI 153
 QY 121 LINNAGVRCPHMTTEDEGFEMQFGVNLGHFLTLNLLDCLKASAPSRINISSLAHVAG 180
 DB 154 LINNAGVRCPHMTTEDEGFEMQFGVNLGHFLTLNLLDCLKASAPSRINISSLAHVAG 213
 QY 181 HIDFDDLNMQTRKNTKAAVCQSKLAIVLFKELSRRIQSGGVTVNA 227
 DB 214 HIDFDDLNMQTRKNTKAAVCQSKLAIVLFKELSRRIQSGGVTVNA 260

RESULT 15
 AAE03357 ID AAE03357 standard; Protein; 273 AA.
 AC AAE03357;
 XX 10-AUG-2001 (first entry)
 DT
 DE Human gene 8 encoded secreted protein fragment, SEQ ID NO:132.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiotensin disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnerary;
 KW cell culture; chemotaxis; food additive; gene therapy;
 KW binding partner identification.

XX
 OS Homo sapiens.
 XX
 PI
 XX
 DR
 DR
 FT Key Location/Qualifiers
 FT MISC-difference 33
 FT /label= Unknown
 FT /note= "Xaa equals any of the naturally occurring
 L-amino acids"
 XX
 PN WO200134800-A1.
 XX
 PD 17-MAY-2001.
 XX
 PF 08-NOV-2000; 2000WO-US30674.
 XX

PR 12-NOV-1999; 99US-0164750.
PR 30-JUN-2000; 2000US-0215128.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Komatsoulis GA, Ebner R, Fiscella M, Wei P;

DR WPI; 2001-329085/34.

XX New nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -

PS Disclosure; Page 509-510; 530pp; English.

CC AAD07705-AAD07759 represent cDNAs corresponding to 19 human secreted
CC protein genes, and AAE03292-AAE03346 represent the proteins they encode.
CC AAE03347-AAE03375 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 19 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays (e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA)). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention.

XX SQ Sequence 273 AA;

Query Match 67.0%; Score 1136; DB 22; Length 273;

Best Local Similarity 97.8%; Pred. No. 1.7e-108;

Matches 222; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSRYLLPLSALGTGAGAVLLKDYVTGACPSKATIPGKTIYVTGANTGIGKQTALELAR 60
DB 34 MSRYLLPLSALGTGAGAVLLKDYVTGACPSKATIPGKTIYVTGANTGIGKQTALELAR 93
QY 61 RGNITLLACRDMKCEAAADIRGETLNHNHNAHLDLASLKSIREFAKIIIEEERVDI 120
DB 94 RGNITLLACRDMKCEAAADIRGETLNHNHNAHLDLASLKSIREFAKIIIEEERVDI 153
QY 121 LINNAGVRCPPHWTTEDEGFQFVNHGFLTLTLLDLKTKASAPSRILNLSLAHVAG 180
DB 154 LINNAGVRCPPHWTTEDEGFQFVNHGFLTLTLLDLKTKASAPSRILNLSLAHVAG 213
QY 181 HIDFDLNMQTRKNTKAYQSKLAIVLFTKELSRRLOGSGVTVA 227
DB 214 HIDFDLNMQTRKNTKAYQSKLAIVLFTKELSRRLOGSGVTVA 260

Search completed: February 9, 2004, 11:42:45
Job time : 54 secs

Printed 9/10/01

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 9, 2004, 11:37:22 ; Search time 57 Seconds
(without alignments)
1498.516 Million cell updates/sec

Title: US-10-007-194A-116
Perfect score: 1695
Sequence: 1 MSRYLLPLPSALGTGAGAAVL.....ESARLYGLEAPSVREOPLPR 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.6	331	4 Q8NBN7	Q8NBN7 homo sapien
2	1443.5	85.2	334	11 Q8CEB7	Q8CEB7 mus musculu
3	1340	79.1	260	4 Q96G88	Q96G88 mus sapien
4	1234	72.8	299	11 Q8CC07	Q8CC07 mus musculu
5	818	47.9	330	5 Q96OC7	Q96OC7 drosophila
6	812	47.9	325	5 Q96OC7	Q96OC7 drosophila
7	810	47.8	330	5 Q9V4Q3	Q9V4Q3 drosophila
8	802.5	47.3	331	5 Q9V4P9	Q9V4P9 drosophila
9	789.5	46.6	296	5 Q9V4Q1	Q9V4Q1 drosophila
10	767.5	45.3	300	5 Q9V4Q2	Q9V4Q2 drosophila
11	745.5	44.0	316	11 Q9D1Y4	Q9D1Y4 mus musculu
12	739.5	43.6	318	4 Q8TC12	Q8TC12 homo sapien
13	739.5	43.6	318	4 Q9Y391	Q9Y391 homo sapien
14	738.5	43.6	316	4 Q96NR8	Q96NR8 homo sapien
15	738.5	43.6	316	11 Q8BYK4	Q8BYK4 mus musculu
16	738.5	43.6	318	4 Q9NRW0	Q9NRW0 homo sapien

17	737.5	43.5	316	4 Q8TAW6	Q8TAW6 homo sapien
18	734.5	43.3	300	11 Q9R1R8	Q9R1R8 mus musculu
19	730.5	43.1	316	11 Q9D0U5	Q9D0U5 mus musculu
20	728.5	43.0	293	11 Q9R1R9	Q9R1R9 mus musculu
21	726	42.8	336	4 Q9HBH5	Q9HBH5 mus sapien
22	722.5	42.6	334	11 Q9ER16	Q9ER16 mus musculu
23	715.5	42.2	355	11 Q9QYF1	Q9QYF1 mus musculu
24	695.5	41.0	296	5 Q8MKN1	Q8MKN1 drosophila
25	686.5	40.5	304	11 Q91WAS	Q91WAS mus musculu
26	668.5	39.4	406	5 Q9W4Q4	Q9W4Q4 drosophila
27	566	33.4	303	16 Q53613	Q53613 mycobacteri
28	561	33.1	290	16 Q8PQH7	Q8PQH7 oceanobacti
29	547.5	32.3	321	10 Q91G18	Q91G18 oryza sativ
30	547	32.3	317	5 Q9VE80	Q9VE80 drosophila
31	534	31.5	338	16 Q53726	Q53726 mycobacteri
32	532	31.4	330	4 Q8N514	Q8N514 homo sapien
33	531.5	31.4	304	16 Q9ZBMS	Q9ZBMS mycobacteri
34	530.5	31.3	327	4 Q96EH7	Q96EH7 homo sapien
35	530	31.3	334	4 Q8WUS4	Q8WUS4 homo sapien
36	525	31.0	287	5 Q8T0J5	Q8T0J5 drosophila
37	523	30.9	310	16 Q912C8	Q912C8 streptomyce
38	515.5	30.4	288	4 Q96GR8	Q96GR8 homo sapien
39	515	30.3	320	10 P81259	P81259 brassica na
40	513.5	30.3	336	16 Q9R999	Q9R999 deinococcus
41	513	30.3	406	5 Q8S246	Q8S246 drosophila
42	511	30.1	331	10 Q8RWJ2	Q8RWJ2 arabidopsis
43	504.5	29.8	260	11 Q8BMX8	Q8BMX8 mus musculu
44	502	29.6	331	10 Q8ICE7	Q8ICE7 arabidopsis
45	497	29.3	321	10 Q80924	Q80924 arabidopsis

ALIGNMENTS

Q8NBN7	PRELIMINARY;	PRT;	331 AA.
Q8NBN7	Q8NBN7		
AC	Q8NBN7		
DT	01-OCT-2002 (TREMBLrel. 22, Created)		
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)		
DE	Hypothetical protein NT2RFP2004966.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	SEQUENCE FROM N.A.		
RA	Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayaishi K., Ishii S.,		
RA	Saito K., Yamamoto U., Wakamatsu A., Nagai T., Nakamura Y.,		
RT	"HRI human cDNA sequencing project."		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
CC	-!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASIS/REDUCTASES		
CC	(SDR) FAMILY		
DR	EMBL; AK075392; BAC11591.1; -.		
DR	Genew; HGNC:1978; RDH13.		
DR	InterPro; IPR002198; ADH_short.		
DR	Pfam; PF00106; adh_short.1.		
DR	PRINTS; PR00080; SDRPFAMILY.		
DR	PROSITE; PS00061; ADH_SHORT.1.		
KW	Hypothetical protein; Oxidoreductase.		
SQ	SEQUENCE 331 AA; 35902 MW; E84A3759D2D274F5 CRC64;		
QY	Query Match	99.6%;	Score 1688; DB 4; Length 331;
QY	Best Local Similarity	99.7%;	Pred No. 8.2e-127;
DB	Matches 330; Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
QY	1 MSRYLLPLPSALGTGAGAAVLKDVTGACPSKATIPGKTVITGATGIGKQALBLAR 60		
DB	1 MSRYLLPLPSALGTGAGAAVLKDVTGACPSKATIPGKTVITGATGIGKQALBLAR 60		
QY	61 RGNITLACRDWEKCAAKDIRGETLNNHVAHHDLASLKSIRFPAKIIIEEERYVDI 120		

```

Db      61 RGNITLACRDMKCEAAAKDREGTINHHVNAHDLASIKSIRFPAKIIIEBERVDI 120
QY      121 LINNAGVRCRPHWTTEDEGEMOGVNHGHELTLLDCLKASAPSRINTSLAHVAG 180
Db      121 LINNAGVRCRPHWTTEDEGEMOGVNHGHELTLLDCLKASAPSRINTSLAHVAG 180
QY      181 HIDEVDLNMOTRKNTKAAVCSQSLAIVFTKELSRLOSGSVTVNALHPGVARTELGRR 240
Db      181 HIDEVDLNMOTRKNTKAAVCSQSLAIVFTKELSRLOSGSVTVNALHPGVARTELGRR 240
QY      241 TGHGSTFSSSTLGPFWLVLVKSPELAQOSTYLAVAEELADVSGKYFDGLKOKAPAPPA 300
Db      241 TGHGSTFSSSTLGPFWLVLVKSPELAQOSTYLAVAEELADVSGKYFDGLKOKAPAPPA 300
QY      301 EDEEVARRLMAESARLVGLEAPSVREOPLPR 331
Db      301 EDEEVARRLMAESARLVGLEAPSVREOPLPR 331

```

RESULT 2

```

Q8CEB7  ID Q8CEB7  PRELIMINARY;  PRT;  334 AA.
AC Q8CEB7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Weekly similar to SD07613P.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL, AK028434; BAC25950.1;
SQ SEQUENCE 334 AA; 36464 MW; 4EBCE1643C1FCEC CRC64;

```

```

Query Match      85.2%; Score 1443.5; DB 11; Length 334;
Best local Similarity 83.6%; Pred. No. 3e-107;
Matches 280; Conservative 21; Mismatches 29; Indels 5; Gaps 2;

```

```

QY      1 MSRYLLPLSALGTGAGAAVLKDYVTGGACPSKATITGKYIVTGANTGIGKOTALBLAR 60
Db      1 MSRYLLPLSALGTGAGAAVLKDYVTGGACPSKATITGKYIVTGANTGIGKOTALBLAR 60
QY      61 RGNITLACRDMKCEAAAKDREGTINHHVNAHDLASIKSIRFPAKIIIEBERVDI 120
Db      61 RGNITLACRDMKCEAAAKDREGTINHHVNAHDLASIKSIRFPAKIIIEBERVDI 120
QY      121 LINNAGVRCRPHWTTEDEGEMOGVNHGHELTLLDCLKASAPSRINTSLAHVAG 180
Db      121 LINNAGVRCRPHWTTEDEGEMOGVNHGHELTLLDCLKASAPSRINTSLAHVAG 180
QY      181 HIDEVDLNMOTRKNTKAAVCSQSLAIVFTKELSRLOSGSVTVNALHPGVARTELGRR 240
Db      181 HIDEVDLNMOTRKNTKAAVCSQSLAIVFTKELSRLOSGSVTVNALHPGVARTELGRR 240
QY      241 TGHGSTFSSSTLGPFWLVLVKSPELAQOSTYLAVAEELADVSGKYFDGLKOKAPAPPA 300
Db      241 TGHGSTFSSSTLGPFWLVLVKSPELAQOSTYLAVAEELADVSGKYFDGLKOKAPAPPA 300
QY      301 EDEEVARRLMAESARLVGLEAPSVREOPLPR 331
Db      301 EDEEVARRLMAESARLVGLEAPSVREOPLPR 331

```

RESULT 3

```

Q86G88  ID Q86G88  PRELIMINARY;  PRT;  260 AA.
AC Q86G88;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
DE Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL, BC009881; AA009881.1; -.
DR InterPro; IPR002198; ADH short.
DR Pfam; PF00106; adh short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KM Hypothetical protein; Oxidoreductase.
SQ SEQUENCE 260 AA; 28794 MW; 12C82526B01A3174 CRC64;

```

Query Match

```

Query Match      79.1%; Score 1340; DB 4; Length 260;
Best local Similarity 100.0%; Pred. No. 4e-99;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      72 MEKCEAAAKDREGTINHHVNAHDLASIKSIRFPAKIIIEBERVDI LINNAGVRCP 131
Db      1 MEKCEAAAKDREGTINHHVNAHDLASIKSIRFPAKIIIEBERVDI LINNAGVRCP 60
QY      132 HWTTEDEGEMOGVNHGHELTLLDCLKASAPSRINTSLAHVAGHIDPDLMNQT 191
Db      61 HWTTEDEGEMOGVNHGHELTLLDCLKASAPSRINTSLAHVAGHIDPDLMNQT 120
QY      192 RRYNTKAAVCSQSLAIVFTKELSRLOSGSVTVNALHPGVARTELGRRHIGSTFEST 251
Db      121 RRYNTKAAVCSQSLAIVFTKELSRLOSGSVTVNALHPGVARTELGRRHIGSTFEST 180
QY      252 TLGPFWLVLVKSPELAQOSTYLAVAEELADVSGKYFDGLKOKAPAPAEDEEVARRLMA 311
Db      181 TLGPFWLVLVKSPELAQOSTYLAVAEELADVSGKYFDGLKOKAPAPAEDEEVARRLMA 240
QY      312 ESARLVGLEAPSVREOPLPR 331
Db      241 ESARLVGLEAPSVREOPLPR 260

```

RESULT 4

```

Q8CC07  ID Q8CC07  PRELIMINARY;  PRT;  299 AA.
AC Q8CC07;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Weekly similar to SD07613P.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).

```

DR EMBL, AK034180; BAC28618.1; --
 SQ SEQUENCE 299 AA; 32367 MW; FA73CF854F25BACA CRC64;
 Query Match 72.8%; Score 1234; DB 11; Length 299;
 Best Local Similarity 74.0%; Pred. No. 1.5e-90;
 Matches 248; Conservative 19; Mismatches 28; Indels 40; Gaps 3;
 QY 1 MSRYLLPSALGTVAAGAVLLKDYVTGACPSKATIPGKTIVITGANTGIGKOTALLAR 60
 DB 1 MSRLPLPSVSTVIGVITLNDYVAGACPSKATIPGKTIVITGANTGIGKOTALLAR 60
 QY 61 RGNVILACRDMKECEAAKDIRGETLNHNHVARHLDLAKSIRREPAKIIIEEERVDI 120
 DB 61 RGNVILACRDMKECEAAKDIRGETLNHNHVARHLDLAKSIRREPAKIIIEEERVDI 113
 QY 121 LINNAGVRCFPHWTTEDEGEFMOFGVNHGPHLLTNLLDKLAKSAPRIINLSLAVAG 180
 DB 114 -----GHFLLTNLLDLKLAKSAPRIINLSLAVAG 145
 QY 181 HIDPDLNMQTRKNTKAAVCSKLAIVFTKELSRRLQSGVTVALHFGVARTELGRH 240
 DB 146 HIDPDLNMQTRKNTKAAVCSKLAIVFTKELSRRLQSGVTVALHFGVARTELGRH 205
 QY 241 TGHSTFSSSTLGPIMVLVSPFELAAQPSSTYLAABELADVSGKFDGLKOKAPAPEA 300
 DB 206 TGMNSAFSGFPLGFFFWLLFESPOLAAQPSSTYLAABELADVSGKFDGLKOKAPAPEA 265
 QY 301 EDEVARRLMAESARLVGLE---APSVREOPLFR 331
 DB 266 EDEVARRLMAESARLVGLE---APSVREOPLFR 299
 RESULT 5
 Q960C7 PRELIMINARY; PRT; 330 AA.
 AC Q960C7;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE SD07613P.
 GN Drosophila melanogaster (Fruit fly).
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Fartan D., Frise E.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 (SDR) FAMILY.
 DR EMBL: AY052124; AAK93548.1; --
 DR FlyBase: FBgn0033205; CG2064.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short.1.
 DR PRINTS: PR00080; SDRFAMILY.
 KM Oxidoreductase.
 SQ SEQUENCE 330 AA; 36591 MW; 1D941F466B2554F5 CRC64;
 Query Match 48.3%; Score 818; DB 5; Length 330;
 Best Local Similarity 53.9%; Pred. No. 3.1e-57;
 Matches 171; Conservative 44; Mismatches 100; Indels 2; Gaps 2;
 QY 5 LLELSALGTVAAGAA-LKDYVTGACPSKATIPGKTIVITGANTGIGKOTALLAR 63
 DB 9 LSPILMMATATGGVIGYFLKEYNQGGFTPTDDETGVFVITGANTGIGKOTALLAR 68
 QY 64 NILACRDMKECEAAKDIRGETLNHNHVARHLDLAKSIRREPAKIIIEEERVDI 123

DB 69 TYVLAACRDMKECEAAKDIRGETLNHNHVARHLDLAKSIRREPAKIIIEEERVDI 128
 QY 124 NAGVRCFPHWTTEDEGEFMOFGVNHGPHLLTNLLDKLAKSAPRIINLSLAVAG 183
 DB 129 NAGVRCFPHWTTEDEGEFMOFGVNHGPHLLTNLLDKLAKSAPRIINLSLAVAG 188
 QY 184 FDDLNMQRKNTKAAVCSKLAIVFTKELSRRLQSGVTVALHFGVARTELGRH 243
 DB 189 VADLN-SEKSYDEGLAYSQSKLANVFTRELARLRQSGVTVALHFGVARTELGRH 247
 QY 244 HGSTFSSSTLGPIMVLVSPFELAAQPSSTYLAABELADVSGKFDGLKOKAPAPEA 303
 DB 248 FQTNLVKFFLKMIMPLTLPKSGAOTSYYAALDPEIKNISGLYFSQCKRPVAPALDD 307
 QY 304 EVARRLMAESARLVGLE 320
 DB 308 KVAKFLMAESEKWTGID 324
 RESULT 6
 Q8MZG9 PRELIMINARY; PRT; 325 AA.
 AC Q8MZG9;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE LP06328P.
 GN Drosophila melanogaster (Fruit fly).
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Drenek D., Fartan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Ceiniker S.;
 RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 (SDR) FAMILY.
 DR EMBL: AY102695; AAM27524.1; --
 DR FlyBase: FBgn0033203; CG2070.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short.1.
 DR PRINTS: PR00080; SDRFAMILY.
 KM Oxidoreductase.
 SQ SEQUENCE 325 AA; 36297 MW; B85F4397AD03COA0 CRC64;
 Query Match 47.9%; Score 812; DB 5; Length 325;
 Best Local Similarity 53.5%; Pred. No. 9.1e-57;
 Matches 167; Conservative 43; Mismatches 96; Indels 6; Gaps 2;
 QY 8 LSAIGTVAAGAVLLKDYVTGACPSKATIPGKTIVITGANTGIGKOTALLAR 67
 DB 18 LSAIG-----IYLNQVYMGQGGFTTKTNTGTGVAIVTGNQGGKKEIVLELARGATVYM 72
 QY 68 ACRDMKECEAAKDIRGETLNHNHVARHLDLAKSIRREPAKIIIEEERVDI 127
 DB 73 ACRDMKECEAAKDIRGETLNHNHVARHLDLAKSIRREPAKIIIEEERVDI 132
 QY 128 MRCFPHWTTEDEGEFMOFGVNHGPHLLTNLLDKLAKSAPRIINLSLAVAG 187
 DB 133 MDCPMLTDEGEFMOFGVNHGPHLLTNLLDKLAKSAPRIINLSLAVAG 192
 QY 188 NMQRKNTKAAVCSKLAIVFTKELSRRLQSGVTVALHFGVARTELGRH 247
 DB 193 N-SEKSYDEGLAYSQSKLANVFTRELARLRQSGVTVALHFGVARTELGRH 251

```
QY 248 FSSITLGIPIFWLVKSPPELLAQPSTYLAVALAEADVSGKYPDGLKOKAPAPAEDEBVAR 307
DB 252 FGLLIAPIIWIFIKTANAGACTTLYALDPSLEKVSGRFYSDCKQKHVGSAAQYDDAQ 311
QY 308 RUMAESALVGL 319
DB 312 FLMAESEKWTGT 323

RESULT 7
QYV4Q3 PRELIMINARY; PRT; 330 AA.
ID QYV4Q3;
AC QYV4Q3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG2064 protein.
GN CG2064 OR BC DNA:SD07613.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morten J.H.C., Blazey R.G., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-R., Yandell M.D., Zhang C.R., McKlos G.L.G.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., McKlos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baer A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Boulton J., Brackstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Durbin K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houson K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA LaSko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murty D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon S., Pollard J., Puri V., Reese M.G.,
RA Palazzolo M., Pittington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spindler A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,
```

```
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jallat M., Kruse D., Li P., Mattei B., Mostrel A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
RA Patel J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Pournavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC Mira S., Crosby M.A., Mathews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnák F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mangall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AB003840; AAF59212.3;
DR FlyBase: FBgn0033205; CG2064.
DR InterPro: IPR002198; ADH short.
DR PRINTS: PRO0080; SDRFAM17.
SQ SEQUENCE 330 AA; 36581 MW; 1D941F46681567F5 CRC64;

Query Match 47.8%; Score 810; DB 5; Length 330;
Best Local Similarity 53.6%; Pred. No. 1.3e-56;
Matches 170; Conservative 44; Mismatches 101; Indels 2; Gaps 2;

QY 5 LLPISALGTVAAGV-LLKDYVTGACPSKATIPKTYIVTGANIGIGKOTLLEIARRGG 63
DB 9 LSPILMPATIGVIGYFLKEWGGKFKDDETKVFLVTGANGIGETALEIARRGG 68
QY 64 NIIACDMKCEAAADIRBETLNHNVAHDLASIKSIREPAKITEEBEVDILIN 123
DB 69 TVYLACRDMRCEKARXDIKETNNQNFSSRLDSSLSIKFTVDGPKKEQKATLIN 128
QY 124 NAGVRCPCWHTTEDEGFQFVNHGHFLTLTLIDKTKASAPSRILNLSLAHVAGHD 183
DB 129 NAGVRCPCWHTTEDEGFQFVNHGHFLTLTLIDKTKASAPSRILNLSLAHVAGHD 188
QY 184 FDDLNWQTRKXNTKAAVCOCKLAIVLFTKELSRRLQSGSVTVNAHPGVARTEIGRTGT 243
DB 189 VADLN-SKRSYDEGLAYOSKLANVLFTEILAKRLEGSGVTVAHPGVDTIELARWAF 247
QY 244 HGSSTFSSITLGIPIFWLVKSPPELLAQPSTYLAVALAEADVSGKYPDGLKOKAPAEDE 303
DB 248 FQTNLVKFEFLKMIWPLKTPKSGAQTSTYALDPELKNISGLYFSDCKEPVAVSGALDD 307
QY 304 EVARLNAESALVGL 320
DB 308 KVAKFLMAESEKWTGLD 324

RESULT 8
QYV4P9 PRELIMINARY; PRT; 331 AA.
ID QYV4P9;
AC QYV4P9; QYV4P8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG30491 protein (AT09608p).
GN CG30495 OR CG17986.
```


OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 RX STRAIN=Berkeley;
 RC MEDLINE=20196006; Pubmed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champs M., Pfeiffer B.D.,
 RA Abriell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,
 RA Jalali M., Kalush F., Kaspen G.H., Ke Z., Kienison J.A., Kechum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Port J., Reese M.G.,
 RA Relauer K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.C.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [12]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barton J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibbegan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Port J., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.C., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [13]
 RP SEQUENCE FROM N.A.
 RA Hirst S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.B., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Seale S.M.J., Smith E., Shu S., Smutnitsk F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT Annotation of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [15]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [16]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,
 RA Champs M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Gaetan H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunco J., Pacle J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003840; AAF59216.3; -;
 DR EMBL; AY122067; AAF52579.1; -;
 DR HSSP; P50162; IAE1.
 DR FlyBase; FBgn0050491; CG30491.
 DR FlyBase; FBgn0050495; CG30495.
 DR InterPro; IPR002198; ADH_Short.
 DR PRINTS; PR00080; SDRFAMILY.
 SQ SEQUENCE 331 AA; 37137 MW; 59B1F6D543BDC4B CRC64;
 Query Match 47.3%; Score 802.5; DB 5; Length 331;
 Best Local Similarity 52.2%; Pred. No. 5,4e-56;
 Matches 164; Conservative 44; Mismatches 105; Indels 1; Gaps 1;
 QY 2 SRVLLPSALGTGAAVAVLVLDVYTGACACPEKATIPGTVIVTGNATGIGKOTALELAR 61
 DB 9 SRAPFLSLFTGTTGLAFPKYDLMQGOFTETMETGTGVFVLTGANTGIGKTVETIARK 68
 QY 62 GGNIIACRDMKCEAAKDIKRGTLNHHVARRHDLASLSIRFPAKTIIEERVDIL 121
 DB 69 GGVVWACRNLKCKCEAREIIVETKNKYVCRCDLASGSIRFVAFKRGSEHLAVL 128
 QY 122 INNAVMRCCHWTTEDEFGFQVNHGHPILNLLDLKAKASAPSRINISLAHVAGH 181
 DB 129 INNAVMRCCHWTTEDEFGFQVNHGHPILNLLDLKAKASAPSRINISLAHVAGH 188
 QY 182 IDDDDLNWCQRKNTKRAAYCOSKATVLTFTSEKRLQSGSVTVNVALPGVARTELGR 241
 DB 189 INTGDLN-SDKSYDEGRAYSQSKLVNLFRELAKRLGTVNTAVALPGVADTEIRIM 247
 QY 242 GIGHSFFSTTLGPILFVLVKSPELAQPSVYLAVAEELAVSGKYPDGLKOKAPAPAB 301
 DB 248 GFFNNFAGLFPVPLPMPFKITPRNGAQTSLVYALDELAKVYTGQYFSDCKLKNAPAPAT 307
 QY 302 DEEVARLVAESAR 315
 DB 308 DTGTAKMLVAVSEK 321
 RESULT 9
 QYV4Q1 PRELIMINARY; PRT; 296 AA.
 ID QYV4Q1
 AC QYV4Q1
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE CG2070 protein.
 GN CG2070.
 OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY; PubMed=10731132;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baker E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck U., Brokstein P., Brotler P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fodor C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaitai M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Pauley J.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupki M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Teclor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 (SDR) FAMILY.
 CC EMBL: AE003840; AAF59214.1; -
 DR HSSP: P50162; IAB1.
 DR Flybase: FBgn0033203; CG2070.
 DR InterPro: IPR002198; ADH short.
 DR Pfam: PF00106; adh short1.1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR Oxidoreductase.
 KM SEQUENCE 296 AA; 3316 MW; 7A0E24590022EA5C CRC64;
 SQ
 Query Match 46.6%; Score 789.5; DB 5; Length 296;
 Best Local Similarity 54.6%; Pred. No. 5.1e-55;
 Matches 160; Conservative 40; Mismatches 92; Indels 1; Gaps 1;

DB 242 GA0TTVALDPSLEKVSGRVSDCKQKHVGAAGQDDAQAFLMAESEKMTGI 294
 RESULT 10
 ID Q9V402 PRELIMINARY; PRT; 300 AA.
 AC Q9V402.
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2002 (TrEMBLrel. 20, Last annotation update)
 DE CG2065 protein (RH23455p).
 GN Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baker E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck U., Brokstein P., Brotler P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fodor C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Foster K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaitai M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Pauley J.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupki M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Teclor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 (SDR) FAMILY.
 CC EMBL: AE003840; AAF59213.1; -
 DR EMBL: AY071710; AAL49332.1; -

DR FlyBase; FBgn003204; CG2065.
DR InterPro; IPR002198; ADH short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
KW Oxidoreductase.
SQ SEQUENCE 300 AA; 33378 MW; 4FE7A159BD57FACA CRC64;

Query Match
Best Local Similarity 45.3%; Score 767.5; DB 5; Length 300;
Matches 153; Conservative 50; Mismatches 94; Indels 1; Gaps 1;

DY QY 27 GGACPSKATIPGKTIVITVGTANTGIGKQTALELARRGGNIILACRDMCKECAAAXDIRGET 86
DB : : : : :
DY Db 3 GGOFTQTDETKVFIVTGANTGIGKETVLEIAKGTVVMACEDMNRCEKAQDIIRET 62
DY QY 87 LNHVNARHLDLASLSIRFAAKIIEBERVDILLNAGVMCRPHWTTEGDGFEMOFGVN 146
DB : : : : :
DY Db 63 NNQIFSRDELSSLESIRKFAAFKEQDKLVLLINNAGVMCRPTLTLDKFEMOLGVN 122
DY QY 147 HLGHFLTLNLDDKLKASAPSRINSLLAHVAGHDFFDNLNQTRKYNTKAAVCOSKLA 206
DB : : : : :
DY Db 123 HMGHFDTLTHLLDVLLKKTAAPSRIINVSVSSVHTQGFIKTADLN-SEKSYSRGAYSQSCLA 181
DY QY 207 IVLFTELKSRRLQSGVTNNALHPGVARTELGRHTGIHGTFSTTLGPFRFWLLVKSP 266
DB : : : : :
DY Db 182 NVLFTRELAKLGSTGVTNSLPGADVETLSRNWFKLPFPQLLKPLWLVLFTPN 241
DY QY 267 AQPSTYLVAEEADVGSKYFDGLKOKAPAEADEVARLRWABSARLVGLEAPSV 324
DB : : : : :
DY Db 242 GAQTLYAADPALKDVDGLFYDCPKPEVSAAAQQDKTKGFLWAESKWKTVGNSTKV 299

RESULT 11
Q9DIY4 PRELIMINARY; PRT; 316 AA.

ID	Q9DIY4
AC	Q9DIY4
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	A930033NO7RIk protein.
GS	A930033NO7RIk.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	[1] NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Retina;
RR	MEDLINE=21085660; PubMed=11217851;
RX	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nihei K., Kiyosawa H., Kondo S., Yamanaoka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant C., Fieischmann W., Gaasterland T., Gissi C., King B., Kochiya H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli K., Barsch G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarella J., Mommaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.; "Functional annotation of a full-length mouse cDNA collection.";
RL	Nature 409:685-690(2001).
-!	SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
CC	EMBL; AK020927; BAB32258.1; -.
CCR	HSSP; P50162; 1AEI.

DR	FlyBase; FBgn003204; CG2065.
DR	InterPro; IPR002198; ADH short.
DR	Pfam; PF00106; adh_short; 1.
DR	PRINTS; PR00080; SDRFAMILY.
KW	Oxidoreductase.
SQ	SEQUENCE 300 AA; 33378 MW; 4FE7A159BD57FACA CRC64;
	Query Match
	Best Local Similarity 45.3%; Score 767.5; DB 5; Length 300;
	Matches 153; Conservative 50; Mismatches 94; Indels 1; Gaps 1;
QY	27 GGACPSKATIPGKTIVITVGANTGIGKGTALELARRGGNIILACRDMCKCEAAAKDIRGET 86
Db	3 GGOFTQTDETKVFIVTGANTGIGKETVLEIAKRGGTVVMACEDMNRCEKAQDIIRET 62
QY	87 LNHVNARHLDLASLSIRFAFKLIIEBERVDIILNAGVMCRPHWTTEGDFEMQGVN 146
Db	63 NNQNTFSRELDSLESIRKFAAFKEQDKLVLINNAVMHCPTLTLDGFEMQLGVN 122
QY	147 HLGHFLTLNLDDKLKASAPSRRIINSLAHVAHQIDFDLNNQTRKYNTKAYCQSCLA 206
Db	123 HMGHFLLDTHLLDVLLKKTAAPSRIINVSSLVHTQGFIKTADLN-SEKSYSRIGAYSOKLA 181
QY	207 IVLFTKELSRLOQSGVTVNALHPGVARTGLRGHTGIHGTSFTSTLGPFFWLKVSP 266
Db	182 NVLFTRELAKLGSTGVTTNSLPGADVDTLSRNWFELKHPPQALLKPLDWLFKTPRN 241
QY	267 AQPSTYLVAEELADVSGKYFDGLKQKAPAEDEEVARRLWABSARLVGLEAPSV 324
Db	242 GAQTLTYALDPALKDVSGLYFSDCPKPESVAAAQDDKTGFLWAESKWTGVNSIKV 299
	RESULT 11
ID	Q9DIY4 PRELIMINARY; PRT; 316 AA.
AC	Q9DIY4
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	A930032N07Rik protein.
GN	A930032N07Rik.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
[1]	SEQUENCE FROM N.A.
RN	STRAIN=C57BL/6J; TISSUE=Retina;
RC	MEDLINE=21085660; PubMed=11217851;
RR	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA	Aizawa K., Izawa M., Nihei K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Asburner M., Batalov S., Casavant C.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikolaio I., Pesole G., Quackenbush J.,
RA	Schimi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustinch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA	Lynne P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA	Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA	Hayashizaki Y.;
RL	"Functional annotation of a full-length mouse cDNA collection."
CC	Nature 409:685-690(2001).
-!	SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY	
CC	EMBL; AK020927; BAB32258.1; --
CD	HSSP; P50162; 1AEL.
DR	

DR	MGI; MGI:1925224; A930033N07Rik.
DR	InterPro; IPR002198; ADH short.
DR	Pfam; PF00106; adh_short; 1.
DR	PRINTS; PR00080; SDRFAMILY.
KW	Oxidoreductase.
SQ	SEQUENCE 316 AA; 35270 MW; EF9239B9497D253B CRC64;
	Query Match
	Best Local Similarity 44.0%; Score 745.5; DB 11; Length 316;
	Matches 155; Conservative 56; Mismatches 98; Indels 7; Gaps 3;
QY	5 LLPALSALTGTAGAAVLKDVTGGACPSKATIPGKTIVITVGANTGIGKGTALELARCGN 64
Db	6 VLTSLSLIYLPAPSIKRFAGGVCCTNVQIFGKVVIITGANTGIGKETARELARCAR 65
QY	65 IILACRDMCKCEAAAKDIRGTNLHNHVARHLDSLKSIRFAFKLIIEBERVDIILN 124
Db	66 VYIACRDVLKGESAISEIRADTNQSVLRKLDSDTKSIRAERFLAEKKLDIILN 125
QY	125 AGVMCRPHWTTEGDFEMQFGVNHGLHFLTLNLDDKLKASAPSRRIINSLSLAHVAGHIDF 184
Db	126 AGVMCRPHWTTEGDFEMQFGVNHGLHFLTLNLDDKLKASAPSRRIINSLSLAHVAGHIDF 185
QY	185 DDNLWQTRKYNKAAVCQSKLAIVLFTKELSRLOQSGVTVNALHPGVARTGLRGHTGIH 244
Db	186 HDLQGG-KRYCSAFAYCHSKLANLLFTRELAKLGSTGVTTNSLPGADVDTLSRNWFELKHPPQALLKPLDWLFKTPRN 241
QY	245 GSFTSFTLGPFWLVKSPPELAAQSTYLVAEELADVSGKYFDGLKQKAPAEDEEVARRLWABSARLVGLEAPSV 324
Db	240 -NSYLCLLWRFLSPFPFKSTSQAQTSLSHCAEADELPLSGKYFSDCKRMVSSRARNK 298
QY	305 VARRLWAESARLVGLE 320
Db	299 TAERLWNVSCELLGIQ 314
	RESULT 12
ID	Q8TC12 PRELIMINARY; PRT; 318 AA.
AC	Q8TC12
DT	01-JUN-2002 (TrEMBLrel. 21, Created)
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE	Androgen-regulated short-chain dehydrogenase/reductase 1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
[1]	SEQUENCE FROM N.A.
RN	TISSUE=Prostate;
RC	Strausberg R.;
RA	InterPro; IPR002198; ADH short.
RA	Pfam; PF00106; adh_short; 1.
KW	Oxidoreductase.
SQ	SEQUENCE 318 AA; 35376 MW; 5B0C34B552774835 CRC64;
	Query Match
	Best Local Similarity 43.6%; Score 739.5; DB 4; Length 318;
	Matches 155; Conservative 49; Mismatches 89; Indels 13; Gaps 3;
QY	17 AAVLKDVVTGGACPSKATIPGKTIVITVGANTGIGKGTALELARCGNIILACRDMCKE 76
Db	20 AAPQIRKMLSSLGGVCTSTVQLPGKVVVVVTGANTGIGKETAKELAQGARVYLACRDVEKE 79
QY	77 AAAKDIREGTNLHNHVARHLDSLKSIRFAFKLIIEBERVDIILNAGVMCRPHWTTE 136
Db	80 LVAXEIQTNGNQQLVRKLDSDTKSIRAERFLAEKKLHVLLINNAGVMCMYSKTA 139

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2004, 11:26:51 ; Search time 18 Seconds
(without alignments)

864.769 Million cell updates/sec

Title: US-10-007-194A-116

Perfect score: 1695

Sequence: 1 MSRYLLPLSALGTGAGAAVL.....ESARLVGLEAPSVREQLPR 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	364.5	21.5	398	1	POR_DAUCA
2	356	21.0	397	1	Q96d11 daucus caro
3	353.5	20.9	298	1	OXIR_STRAT
4	352	20.8	297	1	OXIR_STRLI
5	346.5	20.4	405	1	POR_ARATH
6	345.5	20.4	399	1	POR_PEA
7	342	20.2	388	1	POR_HORVU
8	339	20.2	388	1	POR_AVEA
9	338.5	20.0	401	1	POR_ARATH
10	338	19.9	398	1	POR_CUCSA
11	336	19.8	322	1	POR_FLEBO
12	331	19.5	313	1	POR_AVEA
13	329	19.4	401	1	POR_ARATH
14	326.5	19.3	395	1	POR_HORVU
15	325	19.2	322	1	POR_SINY3
16	310.5	18.3	458	1	POR_MARPA
17	289	17.1	407	1	YNSI_YEAST
18	242.5	14.3	267	1	Y711_YEAST
19	237.5	14.0	263	1	UCPA_SALTY
20	229.5	13.5	246	1	FABG_THEMA
21	222.5	13.1	263	1	UCPA_ECO57
22	222.5	13.1	263	1	UCPA_ECOLI
23	222	13.1	272	1	DKH1_STRVN
24	218.5	12.9	246	1	FABG_BACSU
25	218.5	12.9	592	1	EPHD_MYCTU
26	215	12.7	247	1	FAGI_SINY3
27	214	12.6	241	1	FABG_RICPR
28	214	12.6	248	1	FABG_AQUAE
29	213.5	12.6	251	1	Y325_THEMA
30	212	12.5	412	1	BLI4_NEUCR
31	210	12.4	320	1	FABG_CUPLA
32	207	12.2	281	1	YL46_BRAJA
33	205	12.1	256	1	RHLG_PSEAE

ALIGNMENTS

RESULT 1

POR_DAUCA

ID POR DAUCA STANDARD; PRT; 398 AA.

AC Q9SDT1; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Prochlorophyllide reductase, chloroplast precursor (EC 1.3.1.33)

DE (PCR) (NADPH-prochlorophyllide oxidoreductase) (POR).

OS Daucus carota (Carrot).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;

OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.

OX NCBI_TaxID=4039;

RN [1]_TaxID=4039;

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Kurodagosun;

RA Sato K., Demura T., Fukuda H.;

RT "Cloning of photosynthesis-related genes and their expression during

somatic embryogenesis in carrot.";

RL Plant Cell Physiol. 38:77-77(1997).

CC -!- FUNCTION: PHOTOTRANSFORMATION OF PROTOCHLOROPHYLLIDE (PCHLIDE) TO

CHLOROPHYLLIDE (CHLIDE).

CC -!- CATALYTIC ACTIVITY: Chlorophyllide A + NADP(+) =

prochlorophyllide + NADPH.

CC -!- PATHWAY: Chlorophyll biosynthesis.

CC -!- SUBCELLULAR LOCATION: Chloroplast.

CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES

(SDR) FAMILY. POR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; AF207691; AAF20949.1; -.

DR HSSP; P14061; 1FDU.

DR InterPro; IPR002198; ADH short.

DR InterPro; IPR005979; Prochl_reduct.

DR Pfam; PF00106; adh_short; 1.

DR TIGRFAMs; TIGR01289; LPOR; 1.

KW Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase; NADP;

Chloroplast; Transit peptide.

FT TRANSIT ? 398

FT CHAIN ? 398

FT PROTOCHLOROPHYLLIDE REDUCTASE.

FT SEQUENCE 398 AA; 43365 MW; C817D57FA7CF27D5 CRC64;

Query Match 21.5%; Score 364.5; DB 1; Length 398;

Best Local Similarity 34.4%; Pred. No. 8.6e-23;

Matches 121; Conservative 48; Mismatches 134; Indels 49; Gaps 15;

QY 8 LSLAGTVAGAAVLKDYVTGGACFSKATIPGKTVIVTGANTGKGTALARRG-GNII 66

55 ISQTGAIRSAVATTPSVNRATGEGKTLRKGSVIITGASSGLGLATATAKALAEATGKWHY 114
67 LACRMEKCAAAKIDIRGETLNHNHVAHLDLALSKSIREFAAKIIEBERVDILINAG 126
115 MACRDFLKAERAAKAGSKPENYTI--MHLDLASLDSVRQVETFRSRERPLDVLVCNA 172
127 V-----MPCPHWTGDFEMQGVNHLGHFLTNLLDKL--KASAPS--RII----- 170
173 VYPTAKEPTYTA--DGFELSVDGTHLGHFLSRLLDLDLNSKDYPSKRLIIVGSITGNTN 231
171 -----NLSSLAHVAGHIDFDLNNQOT-----RKNYTKAAYCOSKLAIVLFTKELSR 216
232 TLAGNVPPKANGLDRLGLAGGL--NGMSSAMIDGAEPDGAAYKDKSKVCNMLTMOEFHR 289
217 RL--QSGGVTVNALHPG--VARTELGRHTGHGTSFTTTLGPIFWLLVKSPELAAQSTYL 274
290 RYHEETGITFASLYPCGATTTGLFRE--HILPLF--RTLFPFPFOKYITKGVVSEARSKRL 345
275 A--VAEELADVSGKFDGLKQAP-----APAEDEEVARRLWAEARLVGLE 319
346 AQWSEPSLTKSGVYVSNMKNKASPFENQLSSEASDVKEKARKWVEVSEKLVGL 397

RESULT 2
ID POR CHLRE STANDARD; PRT; 397 AA.
AC Q39617;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Prochlorophyllide reductase, chloroplast precursor (EC 1.3.1.33)
DE (PCR) (NADPH-prochlorophyllide oxidoreductase) (POR).
GN POR OR LPCR-1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=137C / CC-125;
RX MEDLINE=96197396; PubMed=8616232;
RA Li J., Timko M.P.;
RT "The pc-1 phenotype of Chlamydomonas reinhardtii results from a
RT deletion mutation in the nuclear gene for NADPH:prochlorophyllide
RT oxidoreductase"; 30:15-37(1996).
RL Plant Mol. Biol. 30:15-37(1996).
CC -!- FUNCTION: PHOTOTRANSFORMATION OF PROTOCHLOROPHYLLIDE (PCHLIDE) TO
CC CHLOROPHYLLIDE (CHLIDE).
CC -!- CATALYTIC ACTIVITY: Chlorophyllide A + NADP(+) =
CC Prochlorophyllide + NADPH.
CC -!- PATHWAY: Chlorophyll biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. POR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U36752; AAB04951.1; -.
CC PIR; S71468; S71468.
CC HSSP; P14061; 1FDU.
DR InterPro; IPR002198; ADH short.
DR InterPro; IPR005979; Prochl_reduct.
DR Pfam; PF00106; adh_short; 1.
DR TIGRFAMs; TIGR01289; LPCR; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase; NADP;
KW Chloroplast; Transit peptide.

FT TRANSIT 1 57 CHLOROPLAST (POTENTIAL).
FT CHAIN 58 397 PROTOCHLOROPHYLLIDE REDUCTASE.
SQ SEQUENCE 397 AA; 41871 MW; 4CEFO8F15545754D CRC64;

Query Match 21.0%; Score 356; DB 1; Length 397;
Best Local Similarity 33.1%; Pred. No. 4.3e-22;
Matches 108; Conservative 54; Mismatches 117; Indels 47; Gaps 13;

QY 34 ATTPGKTVITGANTGIGKQTALSLAREG-GNIIILACRDMKECAAAKIDIRGETLNHNHVN 92
DB 81 ATQOKQTAITIGASSGLGLNAKALAAATGEVHVNACRDLFLKAEQAQAKVGMAGSVSI- 139
QY 93 ARHLDLASLKSIREFAAKIIEBERVDILINAGV---MRCPHWTTEGFEQFMQGVNHL 148
DB 140 -LHLDLSLESVRQVQNFPKASGRDLALVCNAAVLPTAKEPRFTA-DGFELSVDGTHL 197
QY 149 GHFLTNLLDKLKASAPSR---II-----NLSSLAHVAGHIDFD 185
DB 198 GHFLTNLLDLDL-K-NAFNKQPRCIIVGSITGNTNTLAGNVPPKANGLDLSGLAAGVPAA 256
QY 186 DLNQTRKYNTKAAAYCOSKLAIVLFTKELSRRL-QSGGVTVNALHPG-VARTELGRHTGI 243
DB 257 NPMMDGQEFNGAKAYKDKSVACVMTVROHQHFDATGTTFASLYPCGATTTGLFRE--- 313
QY 244 HGSFSSSTLGPFWLLVKS---PELAAQPSSTYLAVAEELADVSGKYFD-----GLKQKA 295
DB 314 HVPLEK--TLFPFPFOKYITKGVVSEAEARLAAVSDPKLNKSGAVKWSSTTGTSPDNQ 371
QY 296 PAPEAEDEEVARRLWAEARLVGLEA 321
DB 372 VSEEVADDSKASKLWDISAKLVGLSA 397

RESULT 3
OXIR STRAT STANDARD; PRT; 298 AA.
AC Q03326;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Probable oxidoreductase (EC 1.---).
OS Streptomyces antibioticus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IMRU 3720;
RX MEDLINE=93194813; PubMed=8383668;
RA Yu T.-W., Chen C.W.;
RT "The unstable melC operon of Streptomyces antibioticus is codeleted
RT with a 1n4811-homologous locus";
RL J. Bacteriol. 175:1847-1852(1993).
CC -!- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases
CC (SDR) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M96551; AAA26796.1; -.
CC PIR; A47089; A47089.
DR InterPro; IPR002198; ADH short.
DR Pfam; PF00106; adh_short; 1.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase.
DR NP_BIND 9 33 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 165 165 BY SIMILARITY.
FT SEQUENCE 298 AA; 31415 MW; DD24DD156632E10C CRC64;
SQ

Query Match	20.9%; Score 353.5; DB 1; Length 298;
Best Local Similarity	33.9%; Pred. No. 4.8e-22;
Matches	99; Conservative 42; Mismatches 128; Indels 23; Gaps 6;
QY	36 IPKTVITVCANTGICGTALALEARRGNILLACRDMEKCEAAAKOTRGETLNHHVNARH 95 : : : : : : : : : : : : : : : : : : Db
DB	3 LTGRRAVVTGGASGLCAETVRALAAAAGVTVTRRPLSAEPLVQELAAAGGAGRVTAAE 62 : : : : : : : : : : : : : : : : : :
QY	96 LDIASLKSIREEPAAKIIIEBERVDILINNAQVMRCPHWT--TEGFEMQFGVHLHGHELLT 154 : : : : : : : : : : : : : : : : : :
DB	63 LDLSDPASVESFAR---AWRGPDILVANAGINALPRTTLAPNGWEMQLATNYLGHFALA 119 : : : : : : : : : : : : : : : : : :
QY	155 NLLLDKLKASAPRIINLSLAHVAGHIDDDNLNQWKNTKAAYCQSCLAIVLFVKEL 214 : : : : : : : : : : : : : : : : : :
DB	120 TGLHAALRDAGSARIVVVSSGAHLDAFPDEDAHFAARPDPWPVAYCQSKAADVLFTVG- 178 : : : : : : : : : : : : : : : : : :
QY	215 SRRLQSGGVTVNALHPGVARTELGRHTG-----IHGSTFTSTTLGPFWLVLKSPELAQA 269 : : : : : : : : : : : : : : : : : :
DB	179 ARRWAAADGITVNALPGVIYLTRLQRHYDDETMRAFGVMDQGNNVKPLPY--YKTFPEGAA 236 : : : : : : : : : : : : : : : : : :
QY	270 PSTYLVAABELADYSKYFGLGKKQ-----APAEEADEEVARRIWL 310 : : : : : : : : : : : : : : : : : :
DB	237 TSVLLAASPILKGVTGYRFEDNEQARTVOGEODPGVAAHALDPEAADRIWL 288 : : : : : : : : : : : : : : : : : :

```

RESULT 4
OXIR_STRLI
ID_OXIR_STRLI STANDARD; PRT; 297 AA.
AC R35320;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Probable oxidoreductase (EC 1.-.-.-).
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=66 / 1326; TRANSPOSON-Tn4811;
RX MEDLINE=93077460; PubMed=1332944;
RA Chen C.W., Yu T.-W., Chung H.-M., Chou C.-F.;
RT "Discovery and characterization of a new transposable element,
RT Tn4811, in Streptomyces lividans 66.";
RL J. Bacteriol. 174:7763-7769(1992).
CC -I- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases
(CC (SDR) family.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF0106; adh_short; 1.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase.
FT NE_BIND 9 33 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 164 164 BY SIMILARITY.
SQ SEQUENCE 297 AA; 31248 MW; 9AFC4A0E596EC32 CRC64;

```

Query Match	20.8%	Score 352;	DB 1;	Length 297;
Best Local Similarity	34.5%;	Pred. No. 6.4e-22;		
Matches 102;	Conservative 42;	Mismatches 128;	Indels 24;	Gaps 8;

Qy	36	IPGKTVLTGANTGIGKQTALRLARRGNNILACRDMKECEAAAKDIRGETLNHHVVARH	95
Db	3	LTGRRVAVTVGASGLGAETVRLAAGAEEVTRIHPQSAEPLVQEAAGAG-RVRAEA	61
Qy	96	LDLASLKSIREFAAKIIEESERVDILNNAGWMCPPHW-TEDGFEMQFGVNHILGHFLLT	154
Db	62	LDLSDVASVDSFAR--AWRGPLDILVANAGIMALPTRTLTPYGWEMQLATNYLGHFALA	118
Qy	155	NLLLDKUKASAPSHIINLSLHVAUGHIDPDLDNWQTKYNTKAAQCOSKLAIVLTFKEL	214
Db	119	TGLHAALURDAGSAIVVVVSGAHILGTPDFEDPHFARRPDPWAAAYGNSKTDADVLTVG-	177
Qy	215	SRRLQSGVTNALHPGVARTELGRHTG-----IHGSTFSTTTLGPIFWLLIKVSPELAAQ	269

```

178  ARRWAADGITANALNPGVILTRLQRHVDDTMRAFGVMDQGNVKPLFY--YKTEQGAA 235
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
179  PSTVLAVAEELADVSGKYE-----DG-LKQKAPAEAEDEEVAARRLWAESE 314
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
180  TSVLLAASPLLVNGVTGRYPEDNQEAETVEDGDVQPGVAAHALDPEAADRLMEYGA 291
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 5
PORA_ARATH
ID_PORA_ARATH STANDARD; PRT; 405 AA.
AC Q42536; Q9FK22;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Proteochlorophyllide reductase A, chloroplast precursor (EC 1.3.1.33)
DE (PCR A) NADPH-proteochlorophyllide oxidoreductase A (POR A).
GN GORA OR AF5G54190 OR K18G13.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
OC NCBI_taxID=3702;
      (1)
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CV. Columbia;
RX MEDLINE=9538766; PubMed=7659751;
RA Armstrong G.A., Runge S., Frick G., Sperling U., Apel K.;
RT "Identification of NADPH:protochlorophyllide oxidoreductases A and B:
RT a branched pathway for light-dependent chlorophyll biosynthesis in
RT Arabidopsis thaliana.";
RL Plant Physiol. 108:1505-1517(1995).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. Columbia;
RX MEDLINE=98403884; PubMed=9734815;
RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
RT Sequence features of the regions of 1,367,185 bp covered by 19
RT physically assigned pl and TAC clones.";
RL DNA Res. 5:203-216(1998).
CC -1- FUNCTION: PHOTOTRANSFORMATION OF PROTOCHLOROPHYLLIDE (PCHLIDE). TO
CC CHLOROPHYLLIDE (CHLIDE). PORA MAY ALSO FUNCTION AS A
CC PHOTOPROTECTANT DURING THE TRANSITORY STAGE FROM DARK TO LIGHT.
CC -1- CATALYTIC ACTIVITY: Chlorophyllide A + NADP(+) =
CC protochlorophyllide + NADPH.
CC -1- PATHWAY: Chlorophyll biosynthesis.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN YOUNG SEEDLINGS.
CC -1- DEVELOPMENTAL STAGE: ETIOLATED SEEDLINGS.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. POR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U29699; AAC49043.1; -
CC EMBL; AB013387; BAB11581.1; --
CC HSPG; P14061; IFBU.
CC DR Interpro; IPR005979; Prochl_reduct.
CC DR TIGRFAMs; TIGR01289; LPOR; 1.
CC KW Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase; NADP;
CC Chloroplast; Transit peptide; Multi-gene family.
CC TRANSIT 1 69
CC CHAIN 70 405 PROTOCHLOROPHYLLIDE REDUCTASE A.
CC CONFLICT 37 37 V -> I (IN REF. 1).
CC SEQUENCE 405 AA; 43863 MW; EB82F2CEFO480D2E CRC64;
CC SQ

```

Query Match	20.4%;	Score 346.5;	DB 1;	Length 405;
Best Local Similarity	35.6%;	Pred.No 2.7e-21;		
Matches 119;	Conservative 45;	Mismatches 123;	Indels 47;	Gaps 15;
25	VTGGACPSKATIPGKTVIVTGANTGKQKQPALEARRG-GNIIACRDMEKCEAAAKDIR 83			
79	VTKSSILDRKTLRKGNVVTGASSGLGLATAKALAEATGKWHVIMACRDFLKAERAAQSAG 138			
84	GETLNHHVNARHLDLASLKIRFEAKIIEEERVDILINNAGVMR-----CPHWTTEDGF 139			
139	MPKDSYTV--MHLDLASLDSVRQFVDNFRFAEMPLDVLVCNAAVYQPTANOPTFTAB-GF 195			
140	EMQFGVNHIGHFLLTWLLLDKLKAS-APS-RIINLSLA---HVAGHI-----DFDD 186			
196	ELSVGNHIGHFLLSRLIIIDLLKKNQDYPKRLIIVGISTGNTNTLGNVPKKNLGLRG 255			
187	LNWQTRKYNITKA-----AYCOSKLAIVLFTKXELSRRL-QSGSVTVNALHPG-VA 233			
256	LAGGLNGLNSAMIDGGDFVGAKAYKDSKVCNMLTMQEFHRRPHEDTGITFASLYPGCIA 315			
234	RTBLGRHTGHGTFSTTTGPIFWLLVKS---PELAAQSTYLAVAELADVSGKYFDG 290			
316	TTGLFRE---HIFLF-RTLFPFPQKITKGYVSEAGKELAQVADPSLTSGVYWSW 370			
291	LKQKAP-----APEADEEVARRLWAESARLVGL 319			
371	NKTSASFENQLSQEASDVEKARVWEVSEKLVGL 404			

RESULT 6

OR_PEA STANDARD; PRT; 399 AA.

Q01289; C

01-APR-1993 (Rel. 25, Created)

01-APR-1993 (Rel. 25, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)

Proteochlorophyllide reductase, chloroplast precursor (EC 1.3.1.33)

(PCR) (NADPH-proteochlorophyllide oxidoreductase) (POR).

Pisum sativum (Garden pea).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

eurosid 1; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.

NCBI_TaxID=3888;

[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 65-76.

SFRAIN-CV, Progress No. 9;

MEDLINE=92256817; PubMed=1581573;

Spano A.J., He Z., Michel H., Hunt D.F., Timko M.P.;

"Molecular cloning, nuclear gene structure, and developmental

expression of NADPH: proteochlorophyllide oxidoreductase in pea (Pisum

sativum L.)";

Plant Mol. Biol. 18:967-972(1992).

-!- FUNCTION: PHOTOTRANSFORMATION OF PROTOCHLOROPHYLLIDE (PCHLIDE) TO

CHLOROPHYLLIDE (CHLIDE).

-!- CATALYTIC ACTIVITY: Chlorophyllide A + NADP(+) =

proteochlorophyllide + NADPH.

-!- PATHWAY: Chlorophyll biosynthesis.

-!- SUBCELLULAR LOCATION: Chloroplast.

-!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES

(SDR) FAMILY. POR SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

or send an email to licenses@isb-sib.ch).

EMBL; X63060; CAA44786.1; --


```

CC (SDR) FAMILY. POR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15869; CAA33879.1; -.
CC PIR; S04783; S04783.
CC HSP; P14061; IFDU.
CC InterPro; IPR002198; ADH short.
CC InterPro; IPR005979; Prochl_reduct.
CC Pfam; PF00106; adh short; 1.
CC TIGRFAMs; TIGR01289; LPOR; 1.
CC Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase; NADP;
KW Chloroplast; Transit peptide; Multigene family.
FT CHAIN 1 74 CHLOROPLAST.
FT CHAIN 75 388 PROTOCHLOROPHYLLIDE REDUCTASE A.
SQ SEQUENCE 388 AA; 41181 MW; 86B3BF153D96C129 CRC64;

Query Match 20.2%; Score 342; DB 1; Length 388;
Best Local Similarity 31.9%; Pred. No. 6.1e-21;
Matches 129; Conservative 46; Mismatches 126; Indels 104; Gaps 18;

QY 1 MSRYLLPL-----SALGTGAGAAVLKDYVTGGACPSKA----- 34
DB 1 MALQLLPSTLSVPKGGSMGAVA-----VKDTAAFLGVSSKAKKASLAVRTQVATAPSPV 55
QY 35 -TIPGKT-----VIVTGANTGIGKQTALAEARRG-GNITLACRDMKCEAA 78
DB 56 TTPSGTASSPSGKKTILRQGVWITGASSGLGLAAKALAEATGKWHVVMACRDFLKASKA 115
QY 79 ADIRGETLNHHVNARHLDLASLSIREFAAKIIEEERVDILINNAGMPCPHWT---T 135
DB 116 AK--AAGMADGSYTVWHLDSLDSVRFQVDAFRRAEMPLDLVLCNAAIYRPTARTPTFT 173
QY 136 EDGFEMQFVGNHGLHFLTLNLLDKL-KASAPSRII-----NLS 173
DB 174 ADGHENSVGNHGLHFLARLLMEDLQKSDYPSRMVIVGITSNTLGNVPPKASLG 233
QY 174 SLAHVAG-----HIDFDLNNWQTKYNTKAAVCOSKLAIVFTKELSRRL-QSGSV 223
DB 234 DLRLAGLGLSGASGAMIDGDE-----SPDGAKAYKDSKVCNMLTMQEFHRRYHEETGI 287
QY 224 TVNALHPG-VARTELGRHTGIHGSTTLPFIWLLV-KSPELAAQFSTYLA--VAEE 279
DB 288 TFSLLYFGCIATTGLFRE---HIPLF--RTLFPFPKFKVTKGVSEASGKRLAQWVAEP 342
QY 280 LADVSGKYFDGLKQAP-----APEAEDEVARRLMAESARLVGL 319
DB 343 VLTGSGVYWSNWKDSASFENQLSQEASDPEKARKVWELSEKLVGL 387

RESULT 8
PORA_WHEAT STANDARD; PRT; 388 AA.
AC Q41578;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protochlorophyllide reductase A, chloroplast precursor (EC 1.3.1.33)
DE (FCR A) (NADPH-protochlorophyllide oxidoreductase A) (POR A).
GN PORA.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Shoot;
RX MEDLINE=94071829; PubMed=8250847;
RA Teakle G.R., Griffiths W.T.;
RT "Cloning, characterization and import studies on protochlorophyllide
RL Biochem. J. 296:225-230(1993).
CC -!- FUNCTION: PHOTOTRANSFORMATION OF PROTOCHLOROPHYLLIDE (PCHLIDE) TO
CC CHLOROCHLOROPHYLLIDE (CHLOROPHYLL).
CC -!- CATALYTIC ACTIVITY: Chlorophyllide A + NADP(+) =
CC protochlorophyllide + NADPH.
CC -!- PATHWAY: Chlorophyll biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. POR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76532; CAA54042.1; -.
CC PIR; S39394; S39394.
CC HSP; P14061; IFDU.
CC InterPro; IPR002198; ADH short.
CC InterPro; IPR005979; Prochl_reduct.
CC Pfam; PF00106; adh short; 1.
CC TIGRFAMs; TIGR01289; LPOR; 1.
CC Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase; NADP;
KW Chloroplast; Transit peptide; Multigene family.
FT CHAIN 1 74 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 75 388 PROTOCHLOROPHYLLIDE REDUCTASE A.
SQ SEQUENCE 388 AA; 41156 MW; 86B9DB96BD085AB3 CRC64;

Query Match 20.0%; Score 339; DB 1; Length 388;
Best Local Similarity 33.9%; Pred. No. 1.1e-20;
Matches 121; Conservative 45; Mismatches 123; Indels 68; Gaps 16;

QY 9 SALGTGAGAAVLKDYVTGGACPS-KATIPKTVIVTGANTGIGKQTALAEARRG-GNII 66
DB 53 SSVTTPSGSAT-----AKPSGKKTILRQGVWITGASSGLGLAAKALAEATGKWHV 103
QY 67 LACRDMKCEAAKDIRGETLNHHVNARHLDLASLSIREFAAKIIEEERVDILINNAG 126
DB 104 MACRDFLKASKAAK--AAGMADGSYTVWHLDSLDSVRFQVDAFRRAEMPLDLVLCNA 161
QY 127 VMRCPEWT---TEDGFEMQFVGNHGLHFLTLNLLDKL-KASAPSRII----- 170
DB 162 IYRPTARTPTFTADGHENSVGNHGLHFLARLLMEDLQKSDYPSRMVIVGITSNT 221
QY 171 -----NLSLAHVAG-----HIDFDLNNWQTKYNTKAAVCOSKLAIVLFTK 212
DB 222 LAGNVPPKASGLDLRLAGLGLSGASGAMIDGDE-----SPDGAKAYKDSKVCNMLTMQ 275
QY 213 ELSSRL-QSGSVTVNALHPG-VARTELGRHTGIHGSTTLPFIWLLV-KSPELAAQ 269
DB 276 EFHRRYHEETGITFSSLYPGCIATTGLFRE---HIPLF--RTLFPFPKFKVTKGVSEAE 330
QY 270 PSTYLA--VAEELADVSGKYFDGLKQAP-----APEAEDEVARRLMAESARLVGL 319
DB 331 SGKRLAQWVAEPLTSKGVTVSNWKDSASFENQLSQEASDPEKARKVWELSEKLVGL 387

RESULT 9
PORA_ARATH STANDARD; PRT; 401 AA.
AC P21218; Q42537;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protochlorophyllide reductase B, chloroplast precursor (EC 1.3.1.33)

```


(PCR B) (NADPH-protochlorophyllide oxidoreductase B) (POR B).
GN PORB OR A74G27440 OR F27G19.40.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI TaxID=3702;
RL [1]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=95388766; PubMed=7659751;
RA Armstrong G.A., Runge S., Frick G., Sperling U., Apel K.;
RT "Identification of NADPH:protochlorophyllide oxidoreductases A and B:
RT a branched pathway for light-dependent chlorophyll biosynthesis in
RT Arabidopsis thaliana";
RL Plant Physiol. 108:1505-1517 (1995).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=cv. An-2; TISSUE=leaf;
RX MEDLINE=91329695; PubMed=1714319;
RA Benli M., Schuëlz R., Apel K.;
RT "Effect of light on the NADPH-protochlorophyllide oxidoreductase of
RT Arabidopsis thaliana";
RL Plant Mol. Biol. 16:615-625 (1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoerge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delensy M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullach B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Bernaiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLeay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Feilber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedid F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Fumelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedha N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Infed J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shokty N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martensen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana";

Nature 402:769-777 (1999).
-!- FUNCTION: PHOTOTRANSFORMATION OF PROTOCHLOROPHYLLIDE (PCHLIDE) TO
CC CHLOROPHYLLIDE (CHLIDE).
CC -!- CATALYTIC ACTIVITY: Chlorophyllide A + NADP(+) =
CC protochlorophyllide + NADPH.
CC -!- PATHWAY: Chlorophyll biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN SEEDLINGS AND ADULT PLANT.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. POR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U29785; AAC49044.1; --
DR EMBL; AL078467; CAB43876.1; --
DR EMBL; AL161571; CAB81394.1; --
DR PIR; T08936; T08936.
DR HSSP; P14061; 1FDU
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR005979; Prochl_reduct.
DR Pfam; PF0106; adh_short; 1.
DR TIGRfams; TIGR01283; LPOR; 1.
DR Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase; NADP;
KW Chlorosynthesis; Chlorophyll peptide; Multigene family.
FT TRANSIT 1 66 CHLOROPLAST (POTENTIAL).
FT CHAIN 67 401 PROTOCHLOROPHYLLIDE REDUCTASE B.
FT VARIANT 395 395 E -> D (IN STRAIN CV. ANT-2).
SQ SEQUENCE 401 AA; 43359 MW; 0C2132F980AF6CA3 CRC64;

Query Match 20.0%; Score 338.5; DB 1; Length 401;
Best Local Similarity 33.5%; Pred. No. 1.3e-20;
Matches 116; Conservative 38; Mismatches 105; Indels 87; Gaps 15;

QY 33 KATPGKTVITVGANTGIGKGTALRLARRG-GNIIACRDMKCEAAAKDIRGETLNHHV 91
DB 83 KTLRKGNVVVTGASSGLGLATKALATGKWNVIMACEDFLKAERAAKSVGMPKDSYTV 142
QY 92 NARHLDLASKSIREFAAKIIEBERVDILINNAGV---MRCPHWTTDDGFMQGVNH 147
DB 143 --MELDLASLDSVQFVDNFRTERTEPLDVLVCNAAVYFPTAKEPTYSAE-GFELSVA 199
QY 148 LGHFLTNLLDLKLKAS-APS-RII-----NLSLAHVAGHI--- 182
DB 200 LGHFLARLLLDLKLKSDYPSKRLIIVGSITGNTNLGNVPPKANLGLRGLAGLNG 259
QY 183 -----DFDDLNWQTRKYNKAAQCQSLAIVLFTKELSRRL-QSGGVTVNALHPG- 231
DB 260 NSSAMIDGGDFGAK-----AYKDSKVCNMLTMOEFHRRFHRETGVTFASLYPGC 309
QY 232 VARTELGR-HTGIGHST-----SSTLGPFIWLLVKSPELAQAPSTYLA 278
DB 310 IASTGLFREHIPLFLPALFPFPFVKYITKGVSTESKRLAQVVDSPSLTK----- 359
QY 279 ELADVSGKYFGLKQKAP-----APEAEDEVARLWASARLVGL 319
DB 360 -----SGYVSWNNASAFENQLSEASDVAKARKWEISEKLVL 400

RESULT 10
PORA_CUCSA
ID_PORA_CUCSA STANDARD; PRT; 398 AA.
AC Q4249;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protochlorophyllide reductase, chloroplast precursor (EC 1.3.1.33)
DE (PCR) (NADPH-protochlorophyllide oxidoreductase) (POR).

4

Db 299 SEQSDAQAQRMWDLSEKLVGL 321
Search completed: February 9, 2004, 11:43:11
Job time : 21 secs

GN POR OR PCR OR SLR0506.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1] SEQUENCE FROM N.A.
RP MEDLINE=95249551; PubMed=7731978;
RX Suzuki J.Y., Bauer C.E.;
RT "A prokaryotic origin for light-dependent chlorophyll biosynthesis of plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3749-3753(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N., Sugiura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64 to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -!- FUNCTION: PHOTOTRANSFORMATION OF PROTOCHLOROPHYLLIDE (PCHLIDE) TO CHLOROPHYLLIDE (CHLIDE).
CC -!- CATALYTIC ACTIVITY: Chlorophyllide A + NADP(+) = protochlorophyllide + NADPH.
CC -!- PATHWAY: Chlorophyll biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY. POR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; L37783; AAA68281.1; ALT_INIT.
DR EMBL; D64004; BAA10580.1; --
DR PIR; S76636; S76636.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR005979; Prochl_reduct.
DR Pfam; PF00106; adh_short; 1.
DR TIGRFAMs; TIGR01289; LPOR; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase; NADP;
KW Complete proteome.
FT CONFLICT 23 A -> R (IN REF. 1).
SQ SEQUENCE 322 AA; 36061 MW; B805C95E139EA213 CRC64;

Query Match 19.2%; Score 325; DB 1; Length 322;
Best Local Similarity 33.7%; Pred. No. 1.2e-19;
Matches 109; Conservative 58; Mismatches 104; Indels 52; Gaps 17;

QY 40 TVIVTGANTGIGKQTALRARGNIIACRMEKCEAAAKDIRGETLNHHVNAHDLA 99
Db 8 TVIITGASSGVGLYGAALIDKGWHVIMACRNLDKTKVADEL-GFPKDSYTIK-LDLG 65

QY 100 SLKSIREFAAKIIEEERVDILNNAV-----MRCHEWTEDEGFEMQFGVNHIGHFLLTN 155
Db 66 YLDSVRRVFAQFRELGRPLKALVCNAAVFPLLDEPLWSADD-YELSVATNHLGHFLLCN 124

QY 156 LLLDKLKA--SAPSRILNLSLA-----HVAGHIDF--DDL-NWQ-----T 191
Db 125 LLLDLKACPDADKKLIIILGTVTANSKELGKIPAPDPLGNFEGFEAGFKPIAMINN 184

QY 192 RKYNKAAVCQSKALVIFTEKLSREL-QQSGVTNVALHFG-VARTELGRTGHTGHTFS 249
Db 185 KFKSGKAYKDSKLCNMLTRELHRRFHQETGIVFNSLXPGCVADTLPFN---HYSLF- 240

QY 250 STTLGPIFWLLVK-----SPELAQPSITYLAABELADVSGKYFD-GLKQKAP----- 296
Db 241 -RTIFPFQKNTVKGVSQELAGE-RVAMVVAADKFKDSGVHNSWGNRQAGREAFVQEL 298

QY 297 APEADEEVARRLNAESARLVGL 319

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2004, 11:39:21 ; Search time 27 Seconds
(without alignments)
1178.957 Million cell updates/sec

Title: US-10-007-194A-116

Perfect score: 1695

Sequence: 1 MSRYLLPLSALGTGACAAVL.....ESARLVGLEAPSVREQLPLR 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	566	33.4	303	2 E70848	probable oxidoredu
2	534	31.5	311	2 H70829	hypothetical prote
3	531.5	31.4	304	2 T44727	probable oxidoredu
4	515	30.4	320	2 S42551	hypothetical prote
5	513.5	30.3	336	2 H75255	oxidoreductase, sh
6	497	29.3	321	2 T02520	probable oxidoredu
7	487.5	28.8	350	2 T48275	hypothetical prote
8	485.5	28.6	311	2 AD2021	hypothetical prote
9	485	28.6	323	2 E87298	hypothetical prote
10	475.5	28.1	317	2 C70863	hypothetical prote
11	457.5	27.0	316	2 D84206	hypothetical prote
12	456	26.9	317	2 T10561	hypothetical prote
13	451	26.6	309	2 C83017	probable short cha
14	447.5	26.4	332	2 T13447	hypothetical prote
15	438.5	25.9	301	2 AE3195	dehydrogenase Atus
16	437	25.8	333	2 T19314	hypothetical prote
17	428	25.3	339	2 T41570	hypothetical prote
18	414	24.4	311	2 T37155	probable oxidoredu
19	402	23.7	312	2 T37150	probable oxidoredu
20	390.5	23.0	294	2 T05381	hypothetical prote
21	382.5	22.6	204	2 T46363	hypothetical prote
22	381	22.5	400	2 C75365	daunorubicin C-13
23	364	21.5	329	2 AI2023	hypothetical prote
24	362.5	21.4	312	2 D83689	hypothetical prote
25	357	21.1	287	2 S19842	probable oxidoredu
26	356	21.0	397	2 S71468	proteochlorophyllid
27	353.5	20.9	298	2 A47089	probable oxidoredu
28	349.5	20.6	291	2 G97327	short-chain alcoho
29	345.5	20.4	400	2 S20941	proteochlorophyllid

30	345	20.4	339	2 B86906	oxidoreductase yxd
31	342	20.2	388	2 S04783	proteochlorophyllid
32	342	20.2	398	2 T33973	hypothetical prote
33	341.5	20.1	296	2 D82515	ketoreductase XF27
34	339	20.0	388	2 S39394	proteochlorophyllid
35	338.5	20.0	401	2 T08936	proteochlorophyllid
36	338	19.9	398	2 JC4146	proteochlorophyllid
37	336	19.8	322	2 T43931	proteochlorophyllid
38	336	19.8	329	2 T15910	hypothetical prote
39	331	19.5	313	2 S08406	proteochlorophyllid
40	329	19.4	401	2 T00897	proteochlorophyllid
41	328.5	19.4	400	2 S30167	proteochlorophyllid
42	326.5	19.3	395	2 S52285	NADPH2 dehydrogena
43	325	19.2	322	2 S76636	hypothetical prote
44	317.5	18.7	322	2 T04022	hypothetical prote
45	312	18.4	302	2 H70523	hypothetical prote

ALIGNMENTS

RESULT 1

E70848

probable oxidoreductase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: E70848

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, C.; Hamlin, N.; Holroyd, S. Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: E70848

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-303 <COL>

A:Cross-references: GB:AL021428; GB:AL123456; NID:g3261514; PIDN:CAA16249.1; PID:g28087

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV0068

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

F:15-208/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match	33.4%	Score	566	DB	2	Length	303
Best Local Similarity	43.9%	Pred. No.	1.3e-37				
Matches	133	Conservative	47	Mismatches	101	Indels	22
Gaps	8						
QY	34	ATIPDQTGRTAVITGANTGIGKGTAKLELARRGNNIIACRDWKECAAAKDIRGETLNHH	90				
Db	7	ADIPDQTGRTAVITGANTGIGFETAALAAHGAHVLAVRNLDKQKQAAARTEATPGAE	66				
QY	91	VNARHLDLASLKSIRFPAKIIIEBERVDILINNAGVNRCPHWTTEDEGFQGVNHLGH	150				
Db	67	VELQELDLTSLASVRAAAQKSDHQIRIDLLINNAGVWYTPRQTADGFQGVNHLGH	136				
QY	151	FULTNLLDKLKASAPSRRIINLSLAH-VAGHIIFDLDLNWOTRKNTKTAAYCQSKLAIVL	209				
Db	127	FALTGLLIDRLLEPVAGSRVVTISSVGHRIIRAAIHFDLQWE-RRYRVAAYGQAKLANLL	185				
QY	210	FTKELSRRLQSGSVTVN-ALHPGVARTELGHRTGHGTSFTSTLGPFWLLVKSPELAA	268				
Db	186	FTVELQRRLLAPGGTTIAVASHPGVSNTEV-----VRNRPRLVAAILAPDQMAELGA	240				
QY	269	QPSTYLVAABEELADVSGKYF--DGL-----KOKAPAEAEDEFVARRLWAESARLVGL	319				
Db	241	LPI--LRAATDPAVRGGQYFGPDGFGIRGYPKVVAASAQSHDELQRLMAVSEELTGV	298				
QY	320	EAP	322				
Db	299	VVP	301				

```

Query Match          31.4%; Score 531.5; DB 2; Length 304;
Best Local Similarity 43.0%; Pred. No. 7.7e-35;
Matches 132; Conservative 45; Mismatches 101; Indels 29; Gaps 10;

QY      34  ATIP---GKTVITGANTGIGKQTALRLARRGNIILACRDMKCEAAAKDIRGETLNHH 90
DB      7   ADIPDQGRVAVITGANTGLGYQTALAAEHGAHVLAVERNLDKGDAAARITATSAQN 66
QY      91  VNARHLDSLAKSIRREFAAKIIEBERVDILINNAGVMRCPHWTTTDEGFEMQGVNHLGH 150
DB      67  VALQELDLASLESVRAAAKQLRSDYDHIDLLINNAGVMWTPKSTTKDGFELQFGTNNHLGH 126
QY      151  FLTLNLLDKLASAPSRILNISSLAH-VAGHIDFDPLNWMQTRKYNTKAAAYCQSKLAIVL 209
DB      127  FAPTGLLLDRLLPIVGSRVITVSSLSHRFLFADHFNDLQWEC-NYNRVAAYGQSKLANLL 185
QY      210  FTKELSRRLQGGGVTVN-ALHPGVARTELGRHTGIHGSTFSSTTLGPFI---WLLVKSP 264
DB      186  FTYELQRLRLATQTTTIAVAHPGSRTELTRL-----PALIAPIFSVAELFLTQDA 237
QY      265  ELAQPSTYLVAABEELADVSGKYF--DGLKQ-----KAPAPEADEVAR--RLWAESAR 315
DB      238  ATGALPT--LRAATDAAVLGGQYFGPDGFAIRHGPKVVASNGSKSHVDQLRLWAVSEE 295
QY      316  LVGLEAP 322
DB      296  LTGWYP 302

RESULT 4
S42651
hypothetical protein - rape
C:Species: Brassica napus (rape)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C:Accession: S42651
R:Coupe, S.A.; Taylor, J.E.; Isaac, P.G.; Roberts, J.A.
Plant Mol. Biol. 24, 223-227, 1994
A:Title: Characterization of a mRNA that accumulates during development of oil
A:Reference number: S42651; MUID:94154236; PMID:8111020
A:Accession: S42651
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-320 <COU>
A:Cross-references: EMBL:X74225; NID:9456719; PIDN:CAB58175.1; PID:G6065752
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homolog
F:34-235/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match          30.4%; Score 515; DB 2; Length 320;
Best Local Similarity 42.8%; Pred. No. 1.7e-33;
Matches 125; Conservative 50; Mismatches 93; Indels 24; Gaps 8

QY      40  TVIVTANTGTGKOTALRLARRGNIILACRDMKCEAAAKDIRGETLNHHVNAR----H 95
DB      35  TAITTGGTGGIGLEAARVLGNRGARVITASRTK----AANDSKEMILQWYNARIDCLQ 90
QY      96  LDLSLKSIRREFAAKIIEBERVDILINNAGVMRCPHWTTTDEGFEMQGVNHLGHFLTLN 155
DB      91  LDLSISIKSVRSFIHQFLALNVPLNLIINNAGVMFCFQLSDEGIESQFATNHIGHFLTLN 150
QY      156  LLLDLKLKASA-----PSRIINLSLAHVAGH---IDFDLNLWQTRKYNTKAAAYCQSKLAI 207
DB      151  LLLDLQWKSARSBSGIEGRIVNLSIAHTYTYTEGIMFDYIN-DPORYSEKKAYGQSKLAN 209
QY      208  VLFTKELSRRLQGGSV--TVNALHPGVARTELGRHTGIHGSTFSSTTLGPFIWLLVKSP 265
DB      210  LHSNALRGRKQEEGVNITINSVHGLITNLFRISGLGMVLKAMS----PFLWKNIPQ 265
QY      266  LAAQPSTYLVAABEELADVSGKYFDGLKQKAPAPEADEEVARRLWAESARLV 317
DB      266  GAA-ITCYVALHPDLKDVTKGYFADCNVTTPSNFATDTTLADKLMDFSIKLV 316

RESULT 5

```

H75255
oxidoreductase, short-chain dehydrogenase/reductase family - Deinococcus radiodurans (st
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: H75255
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: H75255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <WHI>
A:Cross-references: GB:AE002089; GB:AE000513; NID:G6460427; PIDN:AAF12130.1; PID:G646041
A:Experimental source: strain R1
C:Genetics:
A:Map position: 1
A:Gene: DR2595
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 30.3%; Score 513.5; DB 2; Length 336;
Best Local Similarity 37.9%; Pred. No. 2.4e-33;
Matches 135; Conservative 41; Mismatches 121; Indels 59; Gaps 7;

QY 1 MSRYLLP-----LSALGTGAGAAVLLKDYVTGGACPSKATIPGKTVITVG 45
DB 1 MARRRLPSTHNGNSOSDRPRIISPLAPDAAEVVRG-----VDLKGKTAFTVVG 50
QY 46 AMTGIGKQTALAEARRGNIILACRDMKECEAAKDIRGETLNHHNARHDLASLKSIR 105
DB 51 GASGLTETARALLAGAHVLPVRDRAGERVAAEQST-GGTVELVDLDGLSLASVR 109
QY 106 EFAAKIIEEERVDILINNAVMRCPEHTTDDGFEMQFGVNHGLHFLTLNLLDKLKASA 165
DB 110 RGAETRLQAPRTHILINNAVMATPSQRTVDGFTQGTGNHGLHFLTLRELLPALMAA 169
QY 166 PSRIINLSSLAHVAGHDFDLNMQTKYNTKAAVCOSKLAIVLFTKELSRRLQSGVTV 225
DB 170 PARVVALTSSGHRSDIVMDLNFERRPYDWDAYGOSKNTANLFAVGLTQRYADQGLTA 229
QY 226 NALHPGVARTELGRHGI-----HGSTFSSITLGPFWLLVKSPELAQPSYIL 274
DB 230 NAVHPGGIWTGLQKFVLEPDRRMQWDERG-----TLNPFV---KTPAEGASTSYWA 279
QY 275 AVAEELADVSGKFDGLKQKAP-----APEADESEVARRLWAEARLVG 318
DB 280 ATSPQLYGVGGLFLEDLQHSPTLDESAPNPLFGKPYALDHESARRLWALSEALVG 335

RESULT 6
T02520
Probable oxidoreductase [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F13M22.4
A:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 02-Feb-2001
C:Accession: T02520; A84794
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL data library, June 1998
A:Description: Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.
A:Reference number: Z14677
A:Accession: T02520
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-321 <RCU>
A:Cross-references: EMBL:AC004684; NID:G3236234; PID:G3236237
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Sher, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84794
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-321 <SFO>
A:Cross-references: GB:AE002093; NID:G3236237; PIDN:AAC23625.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g37540; F13M22.4
A:Map position: 2
A:Introns: 40/1; 117/2; 144/1; 195/2; 221/3; 250/1; 265/3
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 29.3%; Score 497; DB 2; Length 321;
Best Local Similarity 40.8%; Pred. No. 4.7e-32;
Matches 119; Conservative 52; Mismatches 97; Indels 24; Gaps 8;

QY 40 TVTIVTGANTGIGKQTALAEARRGNIILACRDMKECEAAKDIRGETLNHHNARV---H 95
DB 35 TAITGTGSGIGLEAARVLAARGARVIAARNPK-----AANESKEMILQMPNARVDYLQ 90
QY 96 LDLASLKSIREFAAKIIEEERVDILINNAVMRCPEHTTDDGFEMQFGVNHGLHFLTLN 155
DB 91 IDVSSIKSVRSFVDQFLANVPLNINNAVMFCFPLKLTEDGIESQFATNHIGHFLTN 150
QY 156 LLLDLKLKASA-----PSRIINLSSLAHVAGH---IDPDLNMQTKYNTKAAVCOSKLAI 207
DB 151 LLLDKMKSTARESGVQGRIVNLSIAHTYTSSEGIKFOGIN-DPAGYSERRAYGQSKLN 209
QY 208 VLFTKELSRRLQSGV---TVNALHPGVARTELGRHTGHSTFTTGLPIFWLLVKSPE 265
DB 210 LLHSNALSRLQEGVNTITNSVHPGLVTNLFYSGFPMKVFRAMTP--LFW---KNIP 264
QY 266 LAAQPTTYLAVALADVSKYPDGLKQKAPAPAEDESEVARRLWAEARLV 317
DB 265 QGAATTCYVALHPDEGVTGKYFGDCNIVAPSKFATNNSLADKLWDFSVELI 316

RESULT 7
T48275
hypothetical protein T22P11.130 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 18-Aug-2000
C:Accession: T48275
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24490
A:Accession: T48275
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-350 <BEV>
A:Experimental source: cultivar Columbia; BAC clone T22P11
C:Genetics:
A:Map position: 5
A:Introns: 40/1; 136/2; 163/1; 214/2; 240/3; 269/1; 284/3
A:Note: T22P11.130
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 28.8%; Score 487.5; DB 2; Length 350;
Best Local Similarity 38.2%; Pred. No. 3e-31;
Matches 118; Conservative 54; Mismatches 98; Indels 39; Gaps 8;

QY 40 TVTIVTGANTGIGKQTALAEARRGNIILACRDMKECEAAKDIRGETLNHHNARHDLA 99
DB 35 TAITGTGSGIGMETARVLKRGHVVIGARNMGAENAKTEILRNANARVTLQLDLUS 94
QY 100 SLKSIREFAAKIIEEERVDILI-----NNAGVMRCPEHTTDDGFE 140
DB 95 SIKSIKAFVREFHALHPLNLLMYTFSLSLIQKFAFAPPPFLANNAGVMFCPQLSEDGIE 154
QY 141 MQFGVNHGLHFLTLNLLDKLKASAPS-----RIINLSSLAHVAGH---IDPDLNMQTR 192
DB 155 LQPATNHIGHFLTLNLLDTMKNTAKTSVGEGRILNVSSVAHVITYQEGIQDSIN-DIC 213

QY 193 KYNTKAYCOSKLAIVLFTKELSRRLQSGV--TVNALHPGVARTELGHTGCI--HGSTF 248
Db 214 SYSDXRAYGSKLANILHANLSROLOEGVYNITANSVHPGLITNLNFQHTALLMRFLKF 273
QY 249 SSTTGLPIFLLVKSPELAQPSYTLVAEELADVSGKYFDGLKQKAPAEAEDEVARR 308
Db 274 FS-----FYLKNIPQGA--TTCYVALHPVSKVGTGYFADCNVETPSKLADETIAQ 326
QY 309 LWAEARLV 317
Db 327 LWFDSVKLI 335

RESULT 8
AD2021
hypothetical protein alr1722 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD2021
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2021
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-311 <KJ>
A:Cross-references: GB:BA000019; PIDN:BA073421.1; PID:g17130812; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1722
C:Superfamily: ribitol dehydrogenase, short-chain alcohol dehydrogenase homology

Query Match 28.6%; Score 485.5; DB 2; Length 311;
Best Local Similarity 40.2%; Pred. No. 3.7e-31;
Matches 117; Conservative 47; Mismatches 112; Indels 15; Gaps 4;

QY 38 GKTIVVTGANTGIGKOTALELARRGGNIILACRDMEXCEAAKDIRGETLNNHVNARHLD 97
Db 16 GLRAIVTSSGLGYETARVLANKQASVITAVNLDKGNKALXILQNKDADVKLMELD 75
QY 98 LASLSIRFPAKIIIEEERVDILNAGVNRCPHWTTEDGFQFQVNHGHFLTLNLL 157
Db 76 LANLAVKFNFAENFRKNYLRDLNINAGVMPYKSTTDGFELQFGTNHLGHFALTQGL 135
QY 158 LDKLASAPRIINLSLAHVAGHIDFDDLNWQTRKYNKTKAAYCOSKLAIVLFTKELSR 217
Db 136 LEPLISTEGSRIVNVSSCAHNMGRKIDFDDLNWQTRKYNKTKAAYCOSKLAIVLFTKELSR 195
QY 218 LQSGV--TVNALHPGVARTELGHTG----IHGTSFSTTLGPIFWLLVKSPELAQ 270
Db 196 LKONGIDTLVTASHPGWTATLQTAGIVKYLNGIVAQDITMGAL-----PTLRAI 248
QY 271 STYLAVALADVSGKYFDGLKQKAPAE--AEDEVARRLWAEARLVGLE 320
Db 249 EAGLKGAIEYFGPNGFMENRGYPIKVESNELSKDQALAKLWVSEKLTQVK 299

RESULT 9
E87298
hypothetical protein CC0398 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: E87298
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; Nierman, W.C.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB87249; MUID:21173698; PMID:11259647

A:Accession: E87298

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <STO>

A:Cross-references: GB:AB005673; NID:g13421559; PIDN:AAK22385.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0398

C:Superfamily: ribitol dehydrogenase, short-chain alcohol dehydrogenase homology

QY 136 EDGFEQMGVGNHGLHFLTNLLDKLKASAPSRINLSSLAHVAGHIDFDLNNWQTRKYN 195
Db 113 ADGFEQFGNHLGHFALTAHLPLRAAQRVSVLSLAARRGRIHFDDLOFE-RSVA 171
QY 196 TKAAYCQSKLAIVLFTKELSRRLQ--SGSVTVNALHPGVARTELGRHTGIGHST----- 247
Db 172 PMTAYCQSKLAVLMFARELDRRSRAAGWGIISNAHPGLTKNLQIAGPSHGDRKPALME 231
QY 248 ---FSSTTLGPIFWLLVKSPELAQPSTYLAELADVSGKYF-----DGLKO- 293
Db 232 RLYKTSWRFAFFLWQELIEGIL---PALY-AAATPQAD--GGAFYGRGRGYEVAGGVREA 286
QY 294 KAPAPAEDEVARRLWAEGARLVLEAPSVR 325
Db 287 KVPAAARNAD-SKELWEVSEQLTGVSYPKSR 317
RESULT 11
D84206
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: D84206
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.
; Leitchauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omar, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: D84206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <STO>
A:Cross-references: GB:AE004437; NID:g10580084; PIDN:AAG19016.1; GSPDB:GN00138
C:Genetics:
A:Gene: yajoi
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
Query Match 27.0%; Score 457.5; DB 2; Length 316;
Best Local Similarity 38.6%; Pred. No. 6.5e-29;
Matches 120; Conservative 43; Mismatches 129; Indels 19; Gaps 8;
QY 23 DVTGACPSKATPGKTVITGANTGIGKQTALEARRGGNIILACRDMKECAAAKDI 82
Db 5 DAWTAALPDQS---GRRVVTGANGSLGFEATAPAGAHVVMACHSTERGEDARDI 61
QY 83 RGTNLHVNARHLDLASKSIREFAAKIIEEERVDILINNAGVMRCPHWTTEGFEHQ 142
Db 62 VAEPLGASLTVHELDLAALDSVAADFADWFTAEDFSLHVLANNAGVMAIPRSETADGFETQ 121
QY 143 FGVNHLGHFLTNLLDKL-KASAPSRINLSSLAHVAGHIDFDLNNWQTRKYNKAAYC 201
Db 122 FGVNHLGHVALTAGLVLTSETGRTVVTQSSGAHRRGRIDFDLQHEA-EYKWEAYS 180
QY 202 QSKLAIVLFTKELSRRLQSGGVTVN--ALHPGVARTELG-RHTGHGTFSTTIGPIFW 258
Db 181 QSKLANLLFAYELDRRLRAASAVTSVACHGYAATNLQLRGPQAGSRLLRLAANA 240
QY 259 LLVKSPELAQPSTYLAELADVSGKYF--DGL-----KQKAPAEAEDEEVARRL 309
Db 241 LVGQSAEQGAWPLLYAATNSID--GGEYIGPGGVNLNRGHPERQOPARSDEDTARLL 298
QY 310 WAESARLVGLE 320
Db 299 WTVSADRTGVD 309
RESULT 12
T10561
C:Species: Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 04-Mar-2000
C:Accession: T10561
R;Bayan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancr
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10561
A:Molecule type: DNA
A:Residues: 1-317 <BEV>
A:Cross-references: EMBL:AL050399; GSPDB:GN00062; ATSP:F25E4.30
A:Experimental source: cultivar Columbia; BAC clone F25E4
C:Genetics:
A:Gene: ATSP:F25E4.30
A:Map position: 4
A:Introns: 36/1; 113/2; 140/1; 191/2; 217/3; 246/1; 261/3
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
Query Match 26.9%; Score 456; DB 2; Length 317;
Best Local Similarity 37.2%; Pred. No. 8.5e-29;
Matches 109; Conservative 56; Mismatches 112; Indels 16; Gaps 5;
QY 38 GKTIVITGANTGIGKQTALEARRGGNIILACRDMKECAAAKDIRGETLNHHVNARHLD 97
Db 29 GLTAIVTGASSGIGETTRVLALRGVHVMAVRNTDSGNQVRDKILKEIPOAKIDVMKLD 88
QY 98 LASLSKIREFAAKIIEEERVDILINNAGVMRCPHWTTEGFEHQGVNHLGHFLTNLL 157
Db 89 LSSMASVRSFASQSLDPLNLINNAGIMACPELSSDNLTELOPATNHLGHFLTNLL 148
QY 158 LDKLKASA----PSRIINSSLAHVAGH---IDFDLNNWQTRKYNKAAQCQSKLAIVL 209
Db 149 LERMKKTASNSRREGRIIVISSEGRFAIRGVQDKINDEAR-YNTLQAYGQSKLGNIL 207
QY 210 FTKELSR--RLQSGVTVNALHPGVARTELGRHTGIGHSTTIGPIFWLLVKSPELA 267
Db 208 HATELARLFKEQGVNITANSIHPGSIWNTLLRY-----HSFINTIGNAVGKYLVKSIPOG 262
QY 268 AQPSTYLAELADVSGKYFDGLKQKAPAEAEDEEVARRLWAESARLVGLE 320
Db 263 AATCTAALHPQAGVSGEVLMDNNNISDPNSQGDKDLAKLWERSLRLTGE 315
RESULT 13
C83017
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83017
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C83017
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <STO>
A:Cross-references: GB:AE004916; GB:AE004091; NID:g9951315; PIDN:AAG08416.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5031
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
Query Match 26.6%; Score 451; DB 2; Length 309;
Best Local Similarity 39.2%; Pred. No. 2.1e-28;
Matches 121; Conservative 37; Mismatches 121; Indels 30; Gaps 9;
QY 32 SKATIP---CKTVITGANTGIGKQTALEARRGGNIILACRDMKECAAAKDIRGETLN 88
Db 5 TRRNIPQLAGRLALVTGANGSLGWQAARTLAGKATVVMACNRREQAERAILDEYPO 64
QY 89 HHVNARHLDLASKSIREFAAKIIEEERVDILINNAGVMRCPHWTTEGFEHQGVNHL 148

Db 65 ARLEIADLADLASIRACAGFRQHRARLDLLFNAGVMFLPLRTRDGFEMQMGTHL 124

Qy 149 GHFLLTNLLDKLASAPSRINLSLAHVAGHIDDDINWQTRKYNTKAAYCOSKLAIV 208

Db 125 GHFALTGLLLSLLAPRVRVGMSTGFNGFQGLPLDDLNAAE-RGNRYLAYCHSKOANL 183

Qy 209 LFTKLSRLQSGVTVNAL--HPGVARTELGRHTGHSSTFSSTTIGPIFWLL----- 260

Db 184 LFSLELQRRAGORGVLLQSLAHPGVAATNL-----QYAAPAMSGSLGR--WAMKVANGA 237

Qy 261 -VKSPELAQPSTYLAVALADSVSGKYFD-----GLKQKAPAP-EAEDDEVARRLW 310

Db 238 FAQSAEMGALPA--LSALTEQRWYGGAYVGPDRWLETRGYPAARIPRNARDLGLAARLW 295

Qy 311 AESARLVGL 319

Db 296 ALSEBLTG 304

RESULT 14

T13447

hypothetical protein T19F6.40 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Mar-2000

C:Accession: T13447; T13449

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew

submitted to the Protein Sequence Database, July 1999

A:Reference number: Z17587

A:Accession: T13447

A:Molecule type: DNA

A:Residues: 1-332 <BEV>

A:Cross-references: EMBL:AL109619; GSPDB:GN00062; ATSP:T19F6.40

A:Experimental source: cultivar Columbia; BAC clone T19F6

C:Genetics:

A:Gene: ATSP:T19F6.40

A:Map position: 4

A:Introns: 41/1; 118/2; 145/1; 197/2; 224/3; 254/1; 269/3

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 26.4%; Score 447.5; DB 2; Length 332;

Best Local Similarity 40.3%; Pred. No. 4.3e-28;

Matches 118; Conservative 47; Mismatches 105; Indels 23; Gaps 9;

Qy 40 TVIVTGANTGIGKGTALAEARRGNIIILACRDMKCEAAAKDIRGETLNHHVNAHLDIA 99

Db 36 TAVITGATSGIGATARVLAKRGARLIFPARNVKAAERKIVSEFPETEIVWKKLDS 95

Qy 100 SLKSIREFAAKIIIEEERVDILINNAGVMRCPHWTEDGFMQFGVNHGLHFLTNLLD 159

Db 96 SIASVRNFVADFESLDLPLNLLINNAGKLAHEHAISEDGIEMTFATNYLGHFLTNLLN 155

Qy 160 KL-----KASAPSRINLSLAH--VAGH-IDFDDLNWQTR-KYNTKAAYCOSKLAIVLF 210

Db 156 KMIQTAEETGVQGRIVNVNTSIGHWFGSLIEYLRLSIQPKCOFQATRAYALSKLANVLH 215

Qy 211 TKELSRLO--GSGVTYNALHPGVARTELGR-HTGHSSTFSSTTIGPIFWL---LVKSP 264

Db 216 TKELSSRLQKIGANVNCVHPGVVTRLTRDREGI-----LTDL--VEFLASKLVKT 267

Qy 265 ELAAQPSYLAVALADSVSGKYFDGLKQKAPAPAEAEDEVARRLWAESARLV 317

Db 268 PQAAATTCYVATNPRLNVNVSQKYFDNCTETPSGLGTNSSEATKLWAASBILV 320

RESULT 15

AE3195

dehydrogenase Atus290 [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasm

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C:Accession: AE3195

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.M.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AE3195

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-301 <KUR>

A:Cross-references: GB:AE008687; PIDN:AA145979.1; PID:gl7743732; GSPDB:GN00188

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atus290

A:Genome: plasmid

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 25.9%; Score 438.5; DB 2; Length 301;

Best Local Similarity 38.2%; Pred. No. 2e-27;

Matches 117; Conservative 52; Mismatches 110; Indels 27; Gaps 11;

Qy 32 SKATIP---GKTIVVTGANTGIGKGTALAEARRGNIIILACRDMKCEAAAKDIRGETLN 88

Db 5 TEANIPNQGRSAVVTGTG-GLGLETALALARAGCDVTIAGRNPEKGSDAVSRIQRAAPH 63

Qy 89 HYNARHLDLASLKSIREFAAKIIEEERVDILINNAGVMRCP-HWTTEDGFMQFGVNH 147

Db 64 VTVSPEKLDLADLTSLAFQRMENDRESLDLLVNNAGIMVPPKQOETRDGFELOFGTNY 123

Qy 148 LGHFLTNLLLDKLSAPSRINLSLAHVAGHIDDDLNWQTRKYNTKAAYCOSKLA 207

Db 124 LGHFALTALHMLPLLLKGTDAVVTVSSVAAPAGKINFADIN-SEKNYHPMRAYSQSKLAC 182

Qy 208 VLFTKELS--RRLQSGSVTVNALHPGVARTELGRHTGHSSTFSSTTIG---PIFWLLVK 262

Db 183 LMFALLEQDRSRAAGVSVSIAHPGVSRITDL-----LHNAPGRNSLQGLARTFLWFLFQ 237

Qy 263 SPELAQPSYLAVALADSVSGKYF--DGLKQ-----KAPA---PEAEDEVARRLWAES 313

Db 238 FVAQGALPOLFSATSKEVK--SGGYGDPRLGETRGHPQPARIPPEALDRVAGKQLWEIS 295

Qy 314 ARLVGL 319

Db 296 QRMTGL 301

Search completed: February 9, 2004, 11:44:47

Job time : 28 secs